UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Advances at the interface: merging information technologies with genomic methodologies

Permalink

https://escholarship.org/uc/item/2542495g

Author

Walker, Faye

Publication Date

2015

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Santa Barbara

Advances at the interface:
merging information technologies with genomic methodologies

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemistry

by

Faye Marguerite Walker

Committee in charge:

Professor H. Tom Soh, Co-Chair

Professor Frederick W. Dahlquist, Co-Chair

Professor R. Daniel Little

Professor Norbert O. Reich

Professor Kevin W. Plaxco

September 2015

The dissertation of Faye Marguerite Walker is approved.

_____
R. Daniel Little

_____
Norbert O. Reich

_____
Frederick W. Dahlquist

_____
Kevin W. Plaxco

_____
H. Tom Soh, Committee Chair

July 2015

Advances at the interface:

merging information technologies with genomic methodologies

Copyright © 2015

by

Faye Marguerite Walker

ACKNOWLEDGEMENTS

This dissertation represents the creative energies of colleagues, friends, and family who have been webbed into the shared experience of my graduate studies. It is beyond a doubt that there are too many names to recognize all the creative energies that have fueled my whirling dervish of work at UCSB, so I will simply focus those without whom my center of gravity would have spun out of control. Prof. Kirill Afonin and Dr. Kareem Ahmad, my primary mentors in the Jaeger and Soh laboratories, respectively, deserve my utmost thanks and gratitude. All past and present members of the Jaeger and Soh groups likewise deserve a round of huzzahs for guidance and wisdom they have imparted. My interactions with the swim, indie music, triathlon, and swing dance communities throughout Santa Barbara and the greater southern California region have been as sweet as the fragrance of roses while supplying a strength that rivals a ridge of granite. More so than anyone else, my family has stood by me through hope and despondency, pleasure and pain, to bring sunshine from rain.

FAYE WALKER

CURRICULUM VITAE OF AN UP-AND-COMER

YOU CAN CONTACT ME HERE:
118 S Glen Annie Road • Goleta, CA 93117
fayewalker@umail.ucsb.edu • (805) 696-3796

EDUCATION

University of California, Santa Barbara, CA (GPA 3.72) 8/2009—9/2015
Doctorate of Philosophy in Chemistry.
Advances at the interface: merging information technologies with genomic methodologies.
H. Tom Soh (chair), R. Daniel Little, Norbert O. Reich, Frederick W. Dahlquist

Swarthmore College, Swarthmore, PA (GPA 3.54) 8/2005—5/2009
HHMI Research Fellow, Class of 1941 Endowed Scholar, Bachelor of Arts with Honors in Chemical Physics.
The study of tenebrionid beetle secretions using solid phase microextraction, gas chromatography-mass spectrometry, and ion mobility spectrometry.
Alison Holliday (chair), Thomas Stephenson, Robert Paley, John Boccio

University of Edinburgh, Edinburgh, UK 8/2008—12/2008
International Visiting Student.

Southern Oregon University, Ashland, OR 9/2008—6/2009
Calculus Scholar.

PEER-REVIEWED MANUSCRIPTS

Walker, F. M., Ahmad, K. M., Eisenstein, M., Soh, H. T. "Transformation of personal computers and mobile phones into genetic detection systems." *Anal Chem* 86, 9236-9241 (2014).

Cho, M., Oh, S. S., Nie, J., Steward, R., Eisenstein, M., Chambers, J., Marth, J. D., **Walker, F. M.**, Thomson, J. A., Soh, H. T. "Quantitative selection and parallel characterization of aptamers," *Proc Nat Acad Sci* 110, 18460-18465 (2013).

Afonin, K. A., Grabow, W. W., **Walker, F. M.**, Bindewald, E., Dobrovolskaia, M. A., Shapiro, B. A., Jaeger, L. "Design and self-assembly of siRNA functionalized nanoparticles for use in automated nanomedicine," *Nat Protoc* 6, 2022-2034 (2011).

Holliday, A. E., **Walker, F. M.**, Brodie, E. D. III, Formica, V. A. "Differences in defensive volatiles of the forked fungus beetle, *Bolitotherus cornutus*, living on two species of fungus," *J Chem Ecol* 35, 1302-1308 (2009).

BOOK PUBLICATIONS

Grabow, W. W., Afonin, K. A., Zakrevsky, P., **Walker, F. M.**, Calkins, E. R., Geary, C., Kasprzak, W., Bindewald, E., Shapiro, B. A., Jaeger, L. "RNA nanotechnology in nanomedicine," *Nanomedicine and Drug Delivery* (2012).

CONFERENCE POSTERS/PRESENTATIONS

Walker, F. M. The Uberification of Healthcare, Graduate Student Showcase Grad Slam, UCSB (Santa Barbara, CA 4/2015).

Walker, F. M., Soh, H. T. The Laboratory of H. Tom Soh: Therapeutic Monitoring, Molecular Diagnostics, and Aptamer Discovery, Dep't of Mechanical Engineering 50th Anniversary Celebration, UCSB (Santa Barbara, CA 10/2014).

Walker, F. M., Ahmad, K. M., Soh, H. T. The P3 System: Molecular Disease Diagnostics Using Mobile Phones and Personal Computers, 24th Annual World Conference on Biosensors, Elsevier (Melbourne, Australia 5/2014).

Walker, F. M. On-site Healthcare by Sleight-of-Hand, Graduate Student Showcase Grad Slam, UCSB (Santa Barbara, CA 4/2014).

Walker, F. M. The Haps with Apps: How Mobile Phones Can Bring Medical Diagnostics to Your Fingertips, Graduate Student Showcase Grad Slam, UCSB (Santa Barbara, CA 4/2013).

Walker, F. M., Holliday, A. "The Study of Tenebrionid Beetle Secretions Using Solid Phase Microextraction, Gas Chromatography-Mass Spectrometry, and Ion Mobility Spectrometry," Sigma Xi Fall Session, Sigma Xi (Swarthmore, PA 5/2009).
Walker, F. M., Schumacher, L., Hilton, C., Holliday, A. "The Effect of Modifier Properties on Chiral Ion Mobility Spectrometry," ISIMS 2008, International Society of Ion Mobility Spectrometry (Ottawa, Canada 7/2008).
Walker, F. M. "Correcting Radiative Lifetime with Quantum Yield in Lead Salt Quantum Dots," Center for Nanoscale Systems Research Experience for Undergradutes, Cornell University (Ithaca, NY 7/2007).
Walker, F. M., Neiswander, B. "Ecological Applications of Wireless Sensor Networks," Photonics Center Research Experience for Undergraduates, Boston University (Boston, MA 7/2006).

RESEARCH EXPERIENCE

Dep't of Mechanical Engineering, Soh Laboratory, UCSB 9/2012—present
Developed platforms for point-of-care testing of infectious diseases by nucleic acid amplification of pathogenic genomic DNA. Performed directed evolution on nucleic acids and assembled biomarker sensors using enzymatic reactions, microfluidic separation, flow cytometry, and cell sorting. Worked in conjunction with industry contacts at Medimmune R&D (Gaithersburg, MD), as well as academic colleagues in the Dep't Molecular, Cellular, and Developmental Biology and creative directors within the College of Engineering.

Dep't of Chemistry and Biochemistry, Jaeger Laboratory, UCSB 9/2009—9/2012
Computationally designed, assembled, and analyzed nanoparticles by means of gel electrophoresis, dynamic light scattering, liquid chromatography, and UV spectroscopy. Operated in concert with contacts at the National Cancer Institute (Bethesda, MD) and supervised two full-time undergraduate student workers. Managed chemical safety compliance.

Dep't of Chemistry and Biochemistry, Swarthmore College 6/2008—6/2009
Pioneered use of custom-built ion mobility spectrometer to separate chiral compounds. Analyzed excretions of forked fungus beetles in collaboration with biologists at the University of Virginia using HS-SPME coupled with GC-MS in an ion trap mass spectrometer.

Center for Nanoscale Systems, Cornell University 6/2007—8/2007
Investigated optical properties of lead semiconductor nanocrystals. Measured absorbance, fluorescence of standard solutions and calculated quantum yields. Synthesized solutions of lead-selenide quantum dots.

Marine Biological Laboratory 1/2007
Researched properties of switchgrass, soybean, and corn as biofuels. Modeled energy outputs of crops using TEM computer application.

Photonics Center, Boston University 6/2006—8/2006
Developed and tested wireless sensor devices for ecological applications. Programmed mote devices with light sensors, calibrated instruments, and simulated environmental processes.

Dep't of Physics and Engineering, Southern Oregon University 6/2004—8/2004
Synthesized two potentially ferroelectric crystals by novel methods. Investigated physical properties of crystals using thermogravimetric, calorimetric, and diffraction analyses.

TEACHING EXPERIENCE

Dep't of Chemistry and Biochemistry, UCSB 8/2009—12/2012

Chem 6BL and 6AL Organic Chemistry Lab Teaching Assistant—Sophomore-level course. Monitored biweekly sessions of four-hour organic chemistry laboratory experiments. Wrote pre-lab lectures, quizzes, and lab report guidelines for each experiment. Instructed students in protocols and theory of reflux, distillation, crystallization, extraction, chromatography, and spectroscopy. Collaborated with instructor in preparing experimental procedures for publication. Proctored and graded midterms and final exams. Assigned grades based on students' ability to synthesize products while using proper techniques in class, in addition to quiz scores and lab report submissions.

Chem 1BL and 1CL General Chemistry Lab Teaching Assistant—Freshman-level course. Instructed two general chemistry laboratory class sections, each meeting once per week for four-hour sessions. Attended weekly organizational meetings with professor on implementing lesson plans. Classes consisted of a quiz, lecture, and experiment based in inorganic, organic, or physical chemistry concepts. Introduced students to the use of voltmeters, spectrophotometers, plus volumetric and heating equipment. Proctored and graded midterms and final exams. Grades were based on in-class efficiency, quiz performance, and lab report quality.

Chem 142/242C Biochemistry Teaching Assistant—Junior-level course.
Involved in the lectures, reviews, and exams of a survey class encompassing the synthetic pathways of nucleic acid and protein biosynthesis. Attended biweekly lectures, held office hours, responded to student inquiries, created potential quiz problems, organized lists of classroom attendance, composed midterms and final exam. Proctored, graded, totaled, and graphically compiled results of midterms and final exam.

Dep't of Physics and Astronomy, Swarthmore College 8/2006—5/2009

Phys 6, 7, 8, and 50 Teaching Assistant—Freshman- and sophomore-level courses.
Attended class lectures, assessed work of physics students, and logged grades. Homework problems were assigned on a weekly basis and dealt with classical mechanics, quantum theory, electromagnetism, and numerical techniques.

VOLUNTEER EXPERIENCE

SciTrek, UCSB 10/2013—present

Module Volunteer—Science outreach program with K-12 students. Visited elementary schools to introduce students firsthand to science through an experimental process over the course of three weeks by designing protocols, carrying out experiments, and facilitating discussion.

Family Ultimate Science Exploration, Santa Barbara 1/2011—5/2011

Workshop Instructor—Middle school extracurricular activities. Prepared chemical experiments, gave hands-on demonstrations, and guided participants through a series of analytical investigations to introduce scientific principles to middle school students and family members.

ADVISEES

Eric Ashley, undergraduate lab intern, UCSB 6/2013—8/2013
Timothy Wu, undergraduate lab intern, UCSB 6/2011—8/2011
Adam Perez, undergraduate lab intern, UCSB 1/2011—6/2011

AWARDS AND FELLOWSHIPS

Doctoral Student Travel Grant, Academic Senate, UCSB (2014)
GSA Excellence in Teaching Award Nominee, Graduate Student Association, UCSB (2011)
Grants-in-Aid of Research, Sigma Xi, The Scientific Research Society (2010)

APPOINTMENTS/MEMBERSHIPS

Recreation Center Governance Board Committee, Representative, UCSB (2013—present)
The Lindy Circle Campus Organization, Officer, UCSB (2013—present)
Women in Science and Engineering, Member, UCSB (2012—present)
Triathlon Team, Member, UCSB (2012—present)
Graduate Students for Diversity in Science, Member (2009—present)
Chemistry Graduate Student Symposium Committee, Member, UCSB (2013—2014)
GSA Excellence in Teaching Award Committee, Member, UCSB (2013)
Graduate Student Association, Dep't of Chemistry and Biochemistry Assembly Representative, UCSB (2012—2013)

MEDIA FEATURES

UCSB in the News, Office of Public Affairs and Communications, UCSB (10/2014)
PBS NOVAnext home page news, PBS.org (10/2014)
PCR Insider home page news, GenomeWeb.com (9/2014)
C&EN home page news, CEN.ACS.org (9/2014)
ACS Editor's Choice article, ACS.org (8/2014)
Mad Science columnist interview, Santa Barbara Sentinel (12/2013)

SKILLS

Business: Microsoft Office, Adobe Photoshop, Adobe Illustrator, Mendeley, EndNote
Computational: MATLAB, Mathematica, RNAfold, RNAcofold, mFold, OligoAnalyzer, Kaleidagraph, Origin
Certified: Biosafety/Bloodborne Pathogens/Medical Waste, Radiation Safety, Laser Safety, Fire Safety, Chemical Lab Safety
Biological Chemistry: FACS, PCR, qPCR, MCA, LAMP, qLAMP, NGS, PAGE, TGGE, DNA extraction, RNA transcription, microarray spotting and scanning, flow cytometry, cell culture, cloning, radiolabeling, immunoassays, pull-downs
Analytical Chemistry: DLS, SLS, HPLC, LC/GC-MS, IMS-MS, sensor and camera technologies, UV/IR/NMR spectroscopy

REFERENCES

Professor H. Tom Soh (Associate Director, Co-Director, Ruth Garland Endowed Chair)
California NanoSystems Institute, Center for Stem Cell Biology and Engineering, Department of Mechanical Engineering, Department of Materials
University of California-Santa Barbara
tsoh@engineering.ucsb.edu • (805) 893-7985

Dr. James B. McGivney (Senior Scientist)
Medimmune LLC
McGivneyJ@MedImmune.com • (301) 398-5106

Dr. Gurpreet Sekhon (Postdoc)
University of California-Santa Barbara
gsekhon@engr.ucsb.edu • (805) 617-8283

ABSTRACT

Advances at the interface:

merging information technologies with genomic methodologies

by

Faye Marguerite Walker

Deoxyribonucleic acid (DNA) is a molecule whose importance towers like a colossus in the sweeping field of biomedicine. The chemical structure of double stranded DNA is itself a helical tower that forms not only the backbone of medical research, but of life itself. Yet, we cannot let DNA be constrained to this role of solid, rigid building block if we wish to utilize its full potential. In its single stranded form, DNA can take on unexpected tertiary shapes that allow it to interact with polymerases, proteins, and organic molecules. This ability gives nucleic acids immense potential for molecular recognition. From monitoring a state of health to identifying toxins in drug development, using DNA as a sensing element can bring valuable information.

Clinical diagnostics have benefited enormously from the sensitivity, specificity, and rapidity of nucleic acid testing (NAT). It can be easy to take blood donor screening, heritable genotyping of newborns, and other standard operating procedures for granted in the developed world. Developing nations with as few as one physician for every 100 people are lacking in the healthcare infrastructure (to say the least) to provide these molecular tests.

The key to unlocking the progressions made by NAT in identifying causative agents of disease comes in the form of a ubiquitous tool: mobile phones. Almost 7B people in the world own cell phones. By combining the optical imaging capabilities and computational powers of mobile phones with a streamlined amplification platform, the ability to detect diseases becomes available to those who could truly reap its benefits.

The synergistic nature of merged technologies is something that extends beyond nucleic acid amplification tests (NAATs) to analysis of recombinant DNA products. Most fields have single-use products that are essential for one purpose but otherwise hidden from view. When brought to the attention of other disciplines, nucleic acid analysis tools such as next generation sequencing (NGS) can segue from providing information on full genomes to identifying highly represented DNA affinity agents from candidate pools. Bringing the systems of particle display, high throughput sequencing (HTS), and *in situ* microarray synthesis to a sequence selection process enhances screening capabilities to the extent that hundreds of sequences can be identified *en masse* as binding to thousands of targets.

TABLE OF CONTENTS

LIST OF FIGURES, TABLES, AND EQUATIONS

Chapter I

Introduction

Since its inception by Francis Crick and James Watson in 1953[1], the three-dimensional model of DNA has remained largely unchanged. The story behind physicist Crick and biologist Watson's landmark discovery is one of applying insider knowledge to an outside field. Crick could explain x-ray diffraction techniques; Watson could predict the orientation of the sugar-phosphate backbone; together, they laid the groundwork for genetics, genomics, and DNA-based studies that are dominating twenty-first century medicine. The message encoded in genomic DNA molecules can explain, on a chemical level, the form and function of human disease. Today, a burgeoning movement in biomedicine has laid the foundation for personalized, preventative, and predictive initiatives[2–4]. Scientific advancements in turning personal devices into monitoring systems are enabling this trend towards rapid, customized diagnostics. The nature of handheld phones and downloadable software makes them an obvious choice in applications for mobile healthcare (mHealth) platforms. Alongside current efforts to adapt bioanalytics for field usage, powerful methods for diagnosing genetic and infectious disease from nucleic acids targets are making pathogen identification and assessment possible on a timeline of minutes or hours. Genomic resources are aiding other industries besides diagnostics, including pharmaceuticals. Approval of therapeutics by regulatory agents becomes easier when the composition of the drug can be understood and decoupled from any possible contaminants. Once more, genome-scale analyses have the ability to enhance product assessment in drug manufacturing pipelines. We will see in the course of this dissertation that when established industries integrate DNA

as a resource in their development of modern assays, the result is a stunning gestalt of analytical force.

A. Thinking small: base by base

Just as GoogleEarth breaks down terabytes of digital mapping data to display street-level and three-dimensional views of a particular city, genetic technologies give clinicians the ability to deconstruct an entire genome of information with single nucleotide resolution. As an increasing number of genomes are sequenced, catalogues of genes can be exploited to serve as targets for clinically useful diagnostic tests. Perhaps the most salient example of NAATs as a powerful clinical tool is their status as the gold standard for detection of *Chlamydia trachomatis* (*C. trachomatis*) and *Neisseria gonorrhoeae* (*N. gonorrhoeae*)[5]. In the case of most infectious diseases, culture or serological techniques are the dominant choice for examining collected specimens. The limitations of these visualizations are markedly apparent: microscopic identification is hugely reliant on the ability of the laboratorian to successfully recognize a pathogen, and culturing methods depend entirely on the time-dependent ability of an organism to propagate on artificial media.

Analysis techniques that operate at the level of the genome have exquisite sensitivity, as well as the ability to do away with the slow turnaround of culturing. Systems based on DNA hybridization technology made early inroads into microbiology through plasmid profiling and bacteriaphage typing[6], but it was the cannonading launch of the polymerase chain reaction (PCR) that simplified and accelerated the outdated arsenal that had been the main weapons of life scientists[7]. The global market for DNA-based diagnostics is expected to be worth $37B by 2018, within which PCR-based amplification tests claim the largest share of $16B[8]. For all the advantages of a well-characterized, developed, and established method

such as PCR, its need for cyclic heating and cooling steps goes hand-in-hand with costly and bulky instrumentation. We sought for a solution to the prohibitive costs that typically impede PCR usage in settings with scant or stretched equipment and found it in a ubiquitous tool: the personal desktop computer (PC).

B. Care for those who will choose, use, and pay the dues

The link between design and function is as strong in the field of medicine as in the market of luxury sports cars. Diagnostic tests and therapeutic treatments are often only suitable for laboratory testing in a clinic. From a global point of view, we can see that eradicating diseases begins with preventative measures to stop the spread of infection and enable treatment—and endemic countries, where resources are often lacking, are where rapid, deployable tests are needed most. Our vision of a platform that uses preexisting products, such as readily available PCs, gets effective results without the need to manufacture or distribute entirely new instruments.

Telemedicine, mHealth, and remote diagnostics all describe ways of leveraging alternative medical technologies to make healthcare more cost-effective and obtainable. A recent report on cell phone activities found that 96% of the world's population (90% being the statistic for developing nations) holds a mobile phone subscription[9]. Penetration of cellular phones is only likely to increase in the coming decades, as the rapid advances in digital technology and the economic forces behind Moore's Law have lowered the overall costs associated with mobile technologies. This steady increase in change is accompanied by refinements in the optical, processive, and connective abilities of cell phones and smart phones[10]. The best aids to point-of-care (POC) diagnostic work may in fact come in the form of a camera phone's imagers and software-enabled computations.

C. The body electric and problematic

The complexity of life has the unfortunate consequence of impeding NAT from the inside out. In biologically derived matrices, the sequential steps of target isolation, amplification, and detection constitute the overall testing process for nucleic acids. Separate steps are a necessity due to the presence of nefarious agents that can obstruct the isolation of nucleic acids from within target cells, degrade the targets beyond recognition, or block the final detection mechanisms from properly measuring the desired molecules[11]. This theme carries through from bodily samples—such as urine, saliva, and blood—to other complex sample types, such as cell lysate. Rather than attempting to perform the primary step of extraction at the POC, numerous research groups have removed this bottleneck by directly detecting nucleic acids in crude sample matrices.

Though it is by no means *sui generis*, the challenges of diagnosing bloodborne pathogens at the POC are greatly in need of developments to identify nucleic acids without extensive pretreatment[12,13]. Along with containing DNA or RNA templates that indicate diseased states, blood is laden with heme compounds, anticoagulants, and EDTA that inhibit DNA polymerase activity[14] or chelate the necessary cofactors[15]. Any alteration that these inhibitory reagents make on the amplification performance will hugely skew attempts to quantify the amount of target DNA present in a sample. Standard techniques to monitor the DNA produced in a reaction are reliant on analyzing a signal in real time. We were driven to investigate whether quantitative techniques could be translated to fieldable systems with minimal technical requirements—both in terms of protocol and equipment.

D. Simplifying assay procedures with simultaneous amplification and detection

NAT methodologies differ widely based on the bodily substance sampled and the amplification method employed. Recent research has made great strides in achieving reliable operation with minimal manipulations, particularly in terms of isothermal amplification techniques. Loop-mediated isothermal amplification (LAMP) has become one of the most widely researched and commercially automated choices for amplification platforms that are amenable to usage at the POC. Early reports in characterizing the tolerance of LAMP for biological substances implied that it was more robust and sensitive than PCR[16–18], and progress in using LAMP for on-site clinical studies is striding forward[19,20]. A critical advancement for clinical LAMP tests, in our reasoning, is the quantitative assessment of pathogenic DNA load. The immediate implication is that therapy can be dispensed in the appropriate amount at the exact time and place where it is needed. To get right to the nexus of whether accurate quantification of blood loads was possible in a field-ready manner, we returned to the use of a ubiquitous tool: a smartphone with a freely available app that could perform continuous measurements and process them in an homage to quantitative real-time thermal cyclers.

E. The spectrum of big data

While this work has been devoted to the shift of diagnostic tools towards availability, there is another massive movement in biomedicine towards the maximization of information. Bioinformatics pushes big data to the extreme with experiments that map whole genomes, transcriptomes, and proteomes[21]. The massively parallel sequencing systems of NGS exemplify the fast-paced frisson of high-throughput technologies. Frederick Sanger developed “first-generation” sequencing in 1977. Shortly after Sanger sequencing was used to complete the human genome project, NGS systems found commercial success with the

introduction of Roche's 454 in 2005 (454 Life Science, Roche Diagnostics, Branford, CT)[22].

The revolutionary reach of NGS extends further than the field of genomics. In the study of nucleic acid affinity reagents, accessing large numbers of sequences is critical for identifying important structure and sequence elements. DNA and RNA molecules that bind with high affinity to a specific target are classified as aptamers. The ability of these single stranded oligonucleotides to fold into complex tertiary structures begets intricate interactions with targets such as peptides, small molecules, and proteins. While the process of generating aptamers by systematic evolution of ligands by exponential enrichment (SELEX)[23,24] has seen extensions and modifications for tackling "difficult" targets[25–27], there remains an inherent bottleneck in characterizing each aptamer's binding capabilities. Our inquiry was whether, by merging the massively parallel genomic investigations of deep sequencing and array-immobilized binding, we could obtain aptamers capable of deconvoluting a portion of an entire proteome.

F. Aptamers as a rising rival to antibodies

A good place to begin talking about the practical application of affinity reagents is with the bellwether of the field: antibodies. The popularity of antibodies as tools for molecular recognition can be seen in the workflow of recombinant protein pharmaceuticals. Many biopharmaceuticals are the product of recombinant proteins expressed in selected host cells. Using host cell lines, which could be prokaryotic (*Escherichia coli*, *E. coli*) or eukaryotic (Chinese hamster ovary, CHO), for biosynthesis results in a complex, crude material full of thousands of proteins that are potential impurities for the final product[28]. For most processing purposes, these destabilizing—and potentially toxic—host cell proteins are

monitored using multi-analyte enzyme-linked immunosorbant assays (ELISAs)[29]. Sandwich-format ELISAs are colorimetric tests that rely on capture antibodies, detection antibodies, and a signal-inducing substrate. The requisite polyclonal antibodies incorporated into ELISAs are generated by injecting animals (typically goats or sheep) with total cell lysate as an immunogen.

When the purpose of quantifying HCPs in recombinant products is to ensure that the US Food and Drug Administration (FDA) standards of less than 100 ppm (w/w) are met, ELISAs are inherently limited. Impediments in using ELISAs come from HCPs that do not generate immune responses in the animal being immunized, low abundance of produced antibody, and nonspecific or oversaturated binding effects[29]. Alongside complications with the final assay are electrostatic and hydrophobic variabilities that exist in HCPs throughout the production process and clinical application[30]. Even more setbacks emerge when considering that polyclonal antibodies themselves take months to develop, vary from batch to batch, and exhibit off-target crossreactivity. We reasoned that aptamers—being thermostable, inexpensive, and commercially available—could bring about a rebirth for HCP testing in recombinant drug development.

Typically, selection is meant to drive a random library towards convergence. Reaching success with a complex biological mixture such as HCP—specifically, over 6,000 proteins are expressed[31] in the case of the CHO genome[32]—depends upon maintaining the molecular diversity of the candidate pool. HTS provides the means of reading a million aptamer candidate sequences after a minimal number of rounds of directed evolution; DNA microarrays act as characterization tools to evaluate aptamer:target binding *en masse* by the tens of thousands. When intertwined, as if in a double helix, these state-of-the-art

approaches can form the basis for a rapid, high-throughput aptamer production pipeline against proteins with a diverse range of charges, hydrophobicities, and structures.

G. Objectives and outline

Today, a burgeoning movement in science has laid the foundations for cross-connectivity. There are two similarly epochal movements that have given rise to the studies presented in this dissertation. Firstly, the decentralization of healthcare has shifted the focus of its technologies towards those that can be widely distributed to peripheral care providers—to the effect that personalized medicine is becoming the way of the future. Secondly, the expansion of big data tools in industry has brought a compulsive reliance on high-throughput methods that can automate the collection and subsequent sorting of volumes of information. Underlying the new paradigms in both of the aforementioned cases is a cross-disciplinary strategy for merging and repurposing available methodologies. Herein, we present three novel platforms engineered for rapid, robust, and user-friendly applications in biomedicine.

Chapter II addresses the unfortunate barrier that has separated the worlds of research development and practical implementation of NAT. It is entirely appropriate to say that each has its own space and place. It is likewise true that each contains aspects of the other, and that efforts to develop technology will only be successful if focused on what the end user wants. We facilitate point-of-care testing by leveraging ubiquitous technology. Mobile healthcare becomes possible—at $1/500^{th}$ of the cost of laboratory instruments—by performing PCR amplification in the heat sink of a standard desktop computer and quantifying the end-point signal with an ordinary camera phone. The PC-PCR-Phone (P3) system is shown to detect genomic DNA (gDNA) from the causative agent of Chagas

disease, *Trypanosoma cruzi*, in whole blood at concentrations fourfold below the average clinical load.

Chapter III expands upon the idea of nucleic acids as agents of change in field-ready diagnostics. Phones once more take a primary position in recording and interpreting the results of a DNA amplification assay. The smartphone real time-LAMP (smaRT-LAMP) platform presented herein is capable of quantitatively measuring *Salmonella enterica* serovar Typhimurium in whole blood collected from septic mice. SmaRT-LAMP requires no manufacture, and distribution is done via a quick app download—all in all, a design based on a marriage of something available and something technical to truly achieve results.

Chapter IV introduces DNA-based affinity reagents as rivals of antibodies in therapeutic and diagnostic applications. In order to truly demonstrate their usefulness in the biomedical field, aptamers must be more than binders with sensitivity and specificity—they must also be functional in biologically relevant environments. The production of therapeutic proteins in a typical CHO cell line can release 1,200-1,400 separate HCP proteins (as visualized by 2-dimensional gel analysis). In order to detect these potentially harmful HCPs by using aptamer reagents, we developed a defined set of DNA aptamers to the convoluted gemisch of HCPs using an innovative selection strategy of particle display screening, microarray characterization, and solid-phase verification of binding. From millions of candidate sequences at the height of our aptamer development pipeline, we whittled the final number down to several dozen unique binders of diverse HCPs that cover 70% of the CHO proteome in their binding capability.

Chapter V summarizes the developments presented in this thesis, evaluates the findings of each project, and proposes future investigations to further the arc of this thematic story of harnessing complementary technologies.

References

[1] J. Watson, F. H. F. Crick, *Nature* **1953**, *171*, 737–8.

[2] S. K. Vashist, O. Mudanyali, E. M. Schneider, R. Zengerle, A. Ozcan, *Anal. Bioanal. Chem.* **2013**, DOI 10.1007/s00216-013-7473-1.

[3] K. a Afonin, W. W. Grabow, F. M. Walker, E. Bindewald, M. a Dobrovolskaia, B. a Shapiro, L. Jaeger, *Nat. Protoc.* **2011**, *6*, 2022–34.

[4] W. W. Grabow, K. A. Afonin, P. Zakrevsky, F. M. Walker, E. R. Calkins, C. Geary, W. Kasprzak, E. Bindewald, B. A. Shapiro, L. Jaeger, in *Nanomedicine Drug Deliv.* (Eds.: M. Sebastian, N. Ninan, A.K. Haghi), Apple Academic Press, **2010**, pp. 208–278.

[5] C. Gaydos, J. Hardick, *Expert Rev. Anti. Infect. Ther.* **2014**, *12*, 657–672.

[6] S. Yang, R. E. Rothman, *Lancet* **2004**, *4*, 337–348.

[7] K. Mullis, F. Faloona, S. Scharf, *Cold Spring Harb.* **1986**, *L1*, 263–273.

[8] BCC, *Diagnostics Market Projected to Grow Nearly $36.5 Billion by 2018*, Wellesley, MA, **2013**.

[9] International Telecommunication Union, *Int. Telecommun. Union* **2014**.

[10] A. Ozcan, *Lab Chip* **2014**, DOI 10.1039/c4lc00010b.

[11] I. G. Wilson, *Appl. Environ. Microbiol.* **1997**, *63*, 3741–3751.

[12] Y. Song, Y.-Y. Huang, X. Liu, X. Zhang, M. Ferrari, L. Qin, *Trends Biotechnol.* **2014**, *32*, 132–9.

[13] A. Afshari, J. Schrenzel, M. Ieven, S. Harbarth, *Crit. Care* **2012**, *16*, 222.

[14] A. Akane, K. Matsubara, D. Ph, H. Nakamura, S. Takahashi, K. Kimura, R. Akane, **1994**, 362–372.

[15] W. A. Al-soud, P. Rådström, *J. Clin. Microbiol.* **2001**, *39*, 485–493.

[16] P. Francois, M. Tangomo, J. Hibbs, E.-J. Bonetti, C. C. Boehme, T. Notomi, M. D. Perkins, J. Schrenzel, *FEMS Immunol. Med. Microbiol.* **2011**, *62*, 41–8.

[17] K. a. Curtis, D. L. Rudolph, S. M. Owen, *J. Virol. Methods* **2008**, *151*, 264–270.

[18] H. Kaneko, T. Kawana, E. Fukushima, T. Suzutani, *J. Biochem. Biophys. Methods* **2007**, *70*, 499–501.

[19] P. Mitashi, E. Hasker, D. M. Ngoyi, P. P. Pyana, V. Lejon, W. Van der Veken, P. Lutumba, P. Büscher, M. Boelaert, S. Deborggraeve, *PLoS Negl. Trop. Dis.* **2013**, *7*, e2504.

[20] Z. K. Njiru, *PLoS Negl. Trop. Dis.* **2012**, *6*, e1572.

[21] S. H. Katsanis, N. Katsanis, *Nat. Rev. Genet.* **2013**, *14*, 415–26.

[22] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu, M. Law, *J. Biomed. Biotechnol.* **2012**, *2012*, DOI 10.1155/2012/251364.

[23] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 183–187.

[24] C. Tuerk, L. Gold, *Science (80-.).* **1990**, *249*, 505–510.

[25] K. M. Ahmad, S. S. Oh, S. Kim, F. M. McClellen, Y. Xiao, H. T. Soh, *PLoS One* **2011**, *6*, e27051.

[26] S. S. Oh, K. M. Ahmad, M. Cho, S. Kim, Y. Xiao, H. T. Soh, *Anal. Chem.* **2011**, *83*, 6883–6889.

[27] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, et al., *PLoS One* **2010**, *5*, DOI 10.1371/journal.pone.0015004.

[28] P. G. Righetti, E. Boschetti, *FEBS J.* **2007**, *274*, 897–905.

[29] J. Zhu-Shimoni, C. Yu, J. Nishihara, R. M. Wong, F. Gunawan, M. Lin, D. Krawitz, P. Liu, W. Sandoval, M. Vanderlaan, *Biotechnol. Bioeng.* **2014**, *111*, 2367–2379.

[30] M. R. Schenauer, G. C. Flynn, A. M. Goetze, *Anal. Biochem.* **2012**, *428*, 150–157.

[31] N. E. Lewis, X. Liu, Y. Li, H. Nagarajan, G. Yerganian, E. O'Brien, A. Bordbar, A. M. Roth, J. Rosenbloom, C. Bian, et al., *Nat. Biotechnol.* **2013**, *31*, 759–65.

[32] X. Xu, H. Nagarajan, N. E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, et al., *Nat. Biotechnol.* **2011**, *29*, 735–741.

Chapter II. Transformation of personal computers and mobile phones into genetic diagnostic systems[1]

Introduction

Although advanced molecular diagnostic technologies for the detection of infectious diseases such as human immunodeficiency virus (HIV), malaria, and tuberculosis[32],[5] are widely available in the developed world, prohibitive costs of equipment and reagents have impeded their adoption in the less developed countries (LDCs) in which these diseases are most prevalent[33–36]. In contrast, access to certain consumer electronics has surged over the past decade in LDCs. For instance, the number of mobile cellular subscribers in the developing world rose by more than 600 million between 2010-2011[37] and desktop computer penetration has dramatically accelerated since the start of the millenium[38].

This trend offers an exciting opportunity for leveraging such tools as a means to affordably improve healthcare in LDCs. For example, the built-in cameras in mobile phones have been adapted as imaging platforms[39,40] for detecting parasites and bacteria[41–43] in blood and other clinically relevant samples[44–47]. However, as these methods are microscopy-based, they can suffer from poor limits of detection and the challenge of differentiating among similar species, subspecies, and strains[48]. Nucleic acid-based genetic tests offer higher sensitivity and exquisite specificity[49,50], and several innovative approaches have been explored to develop low-cost assays and instruments for genetic detection at the point-of-care. For example, Manage *et al.* achieved streamlined detection of BK viruses by performing polymerase chain reaction (PCR) directly in whole blood using

[1] This chapter is adapted with permission from Walker, Faye M.; Ahmad, Kareem M.; Eisenstein, Michael; Soh, H. Tom. *Analytical Chemistry* **2014**, 86, 9326-9241. Copyright 2014 American Chemical Society.

self-contained gel strips[51]. In addition, several groups have demonstrated "sample-in-answer-out" systems that integrate multiple process steps into a monolithic device using microfluidics technology[52–55]. The Landers group pioneered the use of microfluidics for genetic analysis[56], isolating and amplifying nucleic acids directly from buccal swabs and whole blood for clinical examination[57]. Our group has similarly demonstrated direct detection of H1N1 influenza viruses in throat swab samples by integrating magnetic separation with reverse transcriptase PCR (RT-PCR) in a disposable device[58]. In practice, however, the deployment of these systems in low-resource settings is challenging because they often rely on specialized devices and associated instrumentation (*e.g.* pumps, syringes, and detectors) that require skilled technicians for operation.

We have developed an alternate approach to molecular diagnostics that largely eliminates the need for such specialized apparatuses by transforming a desktop PC and a mobile phone—devices that are widely available in LDCs—into a highly sensitive platform for genetic detection of pathogens. Specifically, we use low-cost software tools to convert the PC into a *de facto* thermal cycler for PCR and configure mobile phones as imagers to detect and quantify the resulting PCR amplicons. To our knowledge, this is the first work to perform PCR reactions using a PC. In our PC-PCR-Phone (P3) system, a small volume of patient blood is added directly to a length of disposable tubing that has been preloaded with PCR reagents (**Figure 2.1, step 1**). The tubing is then placed into the heat sink of the central processing unit (CPU), where PCR is performed by using two software programs to precisely manipulate the PC's internal temperature (**Figure 2.1, step 2**). Subsequently, we use a mobile phone camera to image the amplified products, which are quantified according to their fluorescence intensity (**Figure 2.1, step 3**). As a model, we used our P3 system to

detect genomic DNA (gDNA) from *Trypanosoma cruzi* (*T. cruzi*), the parasitic protist responsible for Chagas disease[59,60], which affects over 17 million people worldwide[61,62]. We demonstrate direct detection from blood with a limit of detection of 0.1 fg/μL, which is well below the average parasitic loads found in clinical samples (0.4 fg/μL).

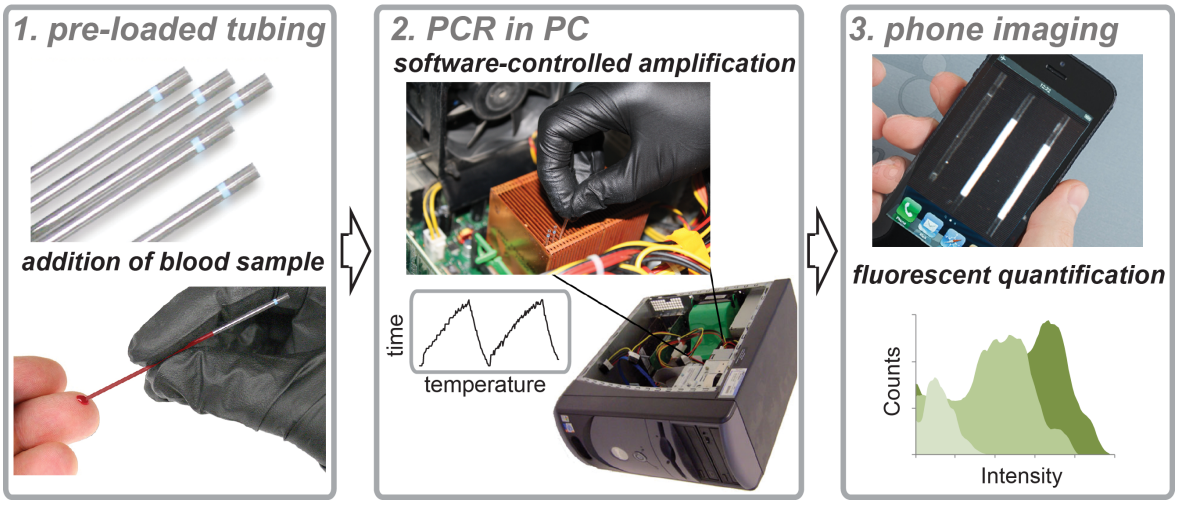

Figure 2.1 P3 assay schematic. First, a small drop of blood obtained via finger prick is added to a length of pre-loaded capillary tubing containing the reagents required for PCR. The tubing is then inserted between the cooling fins on the heat sink in the computer. Commercial software controls CPU usage, cyclically heating and cooling the computer according to a protocol designed to amplify any target DNA present. After thermal cycling, the samples are exposed to light and imaged with a camera phone. By comparing a histogram of the pixel intensities for the patient sample to control samples, the presence or absence of target pathogenic DNA can be determined.

Results and Discussion

A. Efficient amplification of genomic DNA in blood using a PC

PCR amplification of genomic DNA (gDNA) in blood can be hindered by the presence of enzymatic inhibitors naturally found in blood or anticoagulants added after sample collection[63]. To circumvent this problem, we implemented three key modifications to the standard PCR protocol[6] (see Methods). First, we used a step-down (SD)-PCR[64] approach that enables specific and high-yield amplification of long template DNA (*i.e.*, gDNA) with

reduced byproducts[65]. Second, we adopted a two-temperature PCR scheme, consisting of a hot-start followed by alternating hybridization/extension and denaturation steps, that simplifies accurate feedback control of the CPU temperature.

To complete the modified PCR protocol associated with our system, we reduced the denaturation temperature to be lower than the maximum temperature for safe, extended CPU operation (90 °C). We achieved this by adding dimethyl sulfoxide (DMSO) to our PCR mixture, which decreased the melting temperature (T_m) of our primer-template duplex by -0.6 °C per 1% DMSO (**Figure 2.2a**). As an added benefit, the addition of DMSO also improves yield and reduces undesired byproducts[66,67].

B. Software-based thermal cycling in a PC

We installed two software programs that can effectively convert a desktop PC into a PCR thermal cycler. The first program (BurnIn Test) is used to rapidly increase the CPU temperature through intensive computational operations. The second program (SpeedFan) measures the temperature of the CPU in real-time using the built-in thermal sensors common to all CPUs, and controls the cooling fan (See Methods for software details). By running these two programs, the surface temperature of the CPU can be precisely regulated **(Figure 2.2c, black trace)**. However, due to thermal resistance between the heat sink and tubing, the temperature of the blood sample is lower than that of the CPU. In order to correct for this difference, we measured the actual temperature of the sample using thermocouples (**Figure 2.2c, green trace**). These data indicate that the difference in temperature between the CPU and sample does not vary by more than several degrees, and can be accurately predicted with a calibration curve. We established a linear correlation with excellent fit ($R^2 = 0.992$ for

heating and 0.995 for cooling, **Figure 2.2b**) to account for the effect of thermal resistance. Since this thermal resistance should not vary across models of PC, we believe our calibration curve can be used generally and that it is not necessary to calibrate each PC individually.

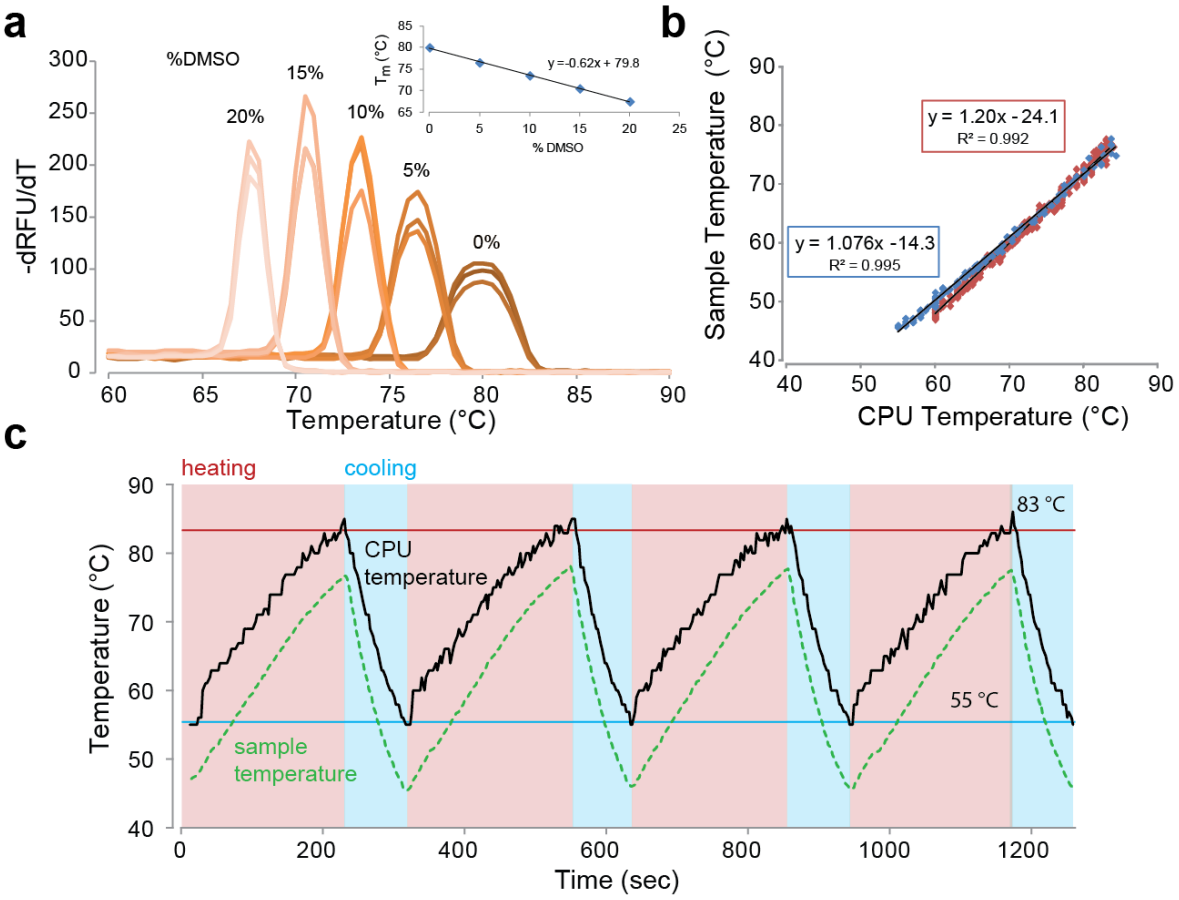

Figure 2.2 With proper temperature calibration and use of additives, DNA is amplified via robust thermal cycling within the PC heat sink. **a.** DMSO reduces melting temperature of double stranded DNA. Amplified DNA was diluted to a range of final DMSO concentrations, and the negative first derivative profiles are shown. Inset: linear correlation of T_m (values from maxima of negative first derivative curves) and % DMSO. **b**. By plotting the software-reported CPU temperature versus the measured sample temperature, we were able to calibrate the sample temperature to the CPU temperature for the heating (red) and cooling steps (blue) separately with best-fit linear regressions. **c**. Temperature traces of the reported CPU temperature (black trace) as recorded by SpeedFan software compared with the measured sample temperature (green trace) as obtained by a thermocouple probe over the course of three cycles. Each two-step PCR cycle started at a temperature of 55 °C for annealing and extension, which was then raised to 83 °C for melting. The vertical bars indicate when the CPU (heating, red bars), and fans (cooling, blue bars) were active.

C. PC-PCR produces single-length amplicons with high yield

Our system is capable of simultaneously performing PCR on up to 29 samples with reproducible yield and no spurious byproducts. To verify that our PCR protocol only generates amplicons of the predicted 100-nt length, we prepared 10 identical samples (20-μL each) and monitored the reaction at every other PCR cycle. Visualization with polyacrylamide gel electrophoresis (PAGE) clearly showed a single product band that matched positive control amplicons obtained with a conventional laboratory thermal cycler (**Figure 2.3a**). We subsequently tested the capacity of our system by performing PCR on 29 identical reactions distributed across the CPU heat sink. We observed minimal variability in amplification (**Figure 2.3b**), both among the various CPU samples and relative to a control amplification performed in a laboratory thermal cycler, as measured by image densitometry following PAGE (C.V. < 4%).

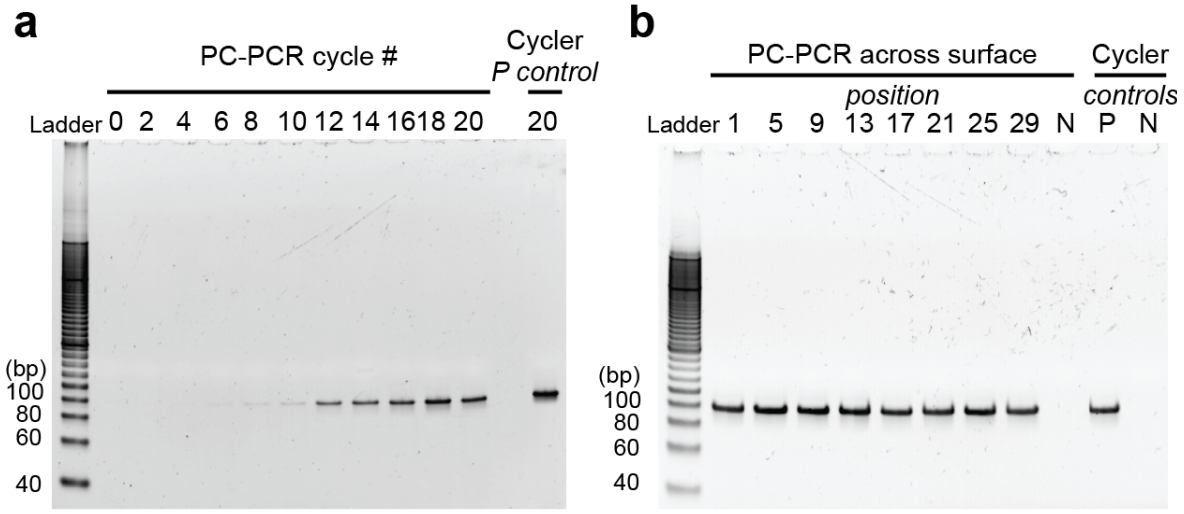

Figure 2.3 PC-PCR amplification is specific and capable of parallelization. **a.** To monitor the progress of PC-based amplification, we amplified 10 identical samples at different positions across the CPU heat sink. One was removed at each cycle, and the resulting products analyzed via gel electrophoresis. **b.** PAGE image showing amplification of a synthetic 100-nt template from reactions performed within the PC heat sink alongside controls. Control reactions were performed by carrying out 20 cycles of PCR on aliquots of the same negative and positive samples in a commercial thermal cycler. Up to 29 reactions may be carried out in parallel.

D. Optical detection with a camera phone for DNA analysis

In order to achieve convenient and quantitative means of readout after amplification, we repurposed a standard camera phone into a quantitative DNA detection platform capable of measuring as little as 9.6 ag of template DNA. Specifically, we outfitted the camera phone with a monochromatic filter to capture fluorescence from a DNA binding dye (see Methods). This dye is present in the PCR mix prior to the reaction, and emits green light (peak wavelength = 520 nm) under UV excitation when complexed with double-stranded DNA amplicons. Using this setup, we were able to clearly differentiate fluorescent signals obtained from PCR reactions performed in a conventional thermal cycler with samples containing as little as 9.6 ag of template DNA relative to template-free negative control samples (**Figure 2.4a**). Moreover, the detection performance of our camera phone system is comparable to that of a laboratory real-time quantitative PCR (qPCR) instrument. Software analysis yielded normalized, mean fluorescence intensity values of our camera phone images (see Methods), which we plotted as a function of template copy number (**Figure 2.4b, black**). We compared these results with the normalized end-point fluorescence values obtained from a qPCR instrument (Bio-Rad iQ5) (**Figure 2.4b, red**). We found that the respective performance of these two platforms correlates very closely, with an $R^2 > 0.99$ (**Figure 2.4c**), and fall within each other's error range at low template copy numbers (< 153 ag).

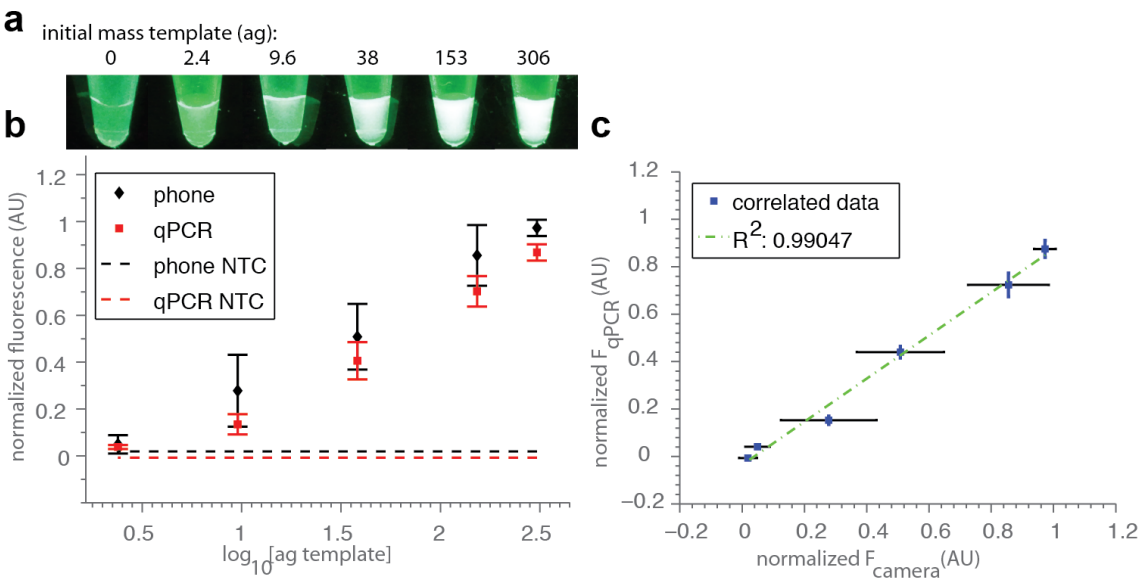

Figure 2.4 Validation of DNA amplicon detection using a mobile phone camera. **a.** Samples containing a range of template molecules were amplified for 30 cycles with EvaGreen, excited by UV transillumination, and imaged with a mobile phone camera and 520 nm filter. **b.** Sensitivity and specificity of camera phone detection of amplified DNA in comparison to qPCR end-point detection. The normalized fluorescence for 10 independent experiments is plotted versus the log of initial copy mass in the reaction. **c**. Best linear fit shown for correlation between mobile phone camera and qPCR end-point fluorescence values for PCR reactions performed on samples initially containing 0 to 306 ag template.

E. Quantitative detection of T. cruzi using the P3 system

The average concentration of *T. cruzi* gDNA in blood samples of patients infected with Chagas disease is 0.4 fg/μL[68]. We found that our P3 system can attain sensitivities below this level from unprocessed whole blood samples. We performed P3 analysis on 20-μL samples containing 1 μL whole blood that had been spiked with 0, 0.1, 1, 10 or 100 fg of *T. cruzi* gDNA. After subtracting the background fluorescence from an empty capillary tube, we could clearly distinguish reactions that had been performed with as little as 0.1 fg/μL gDNA from negative controls (**Figure 2.5a**). PAGE analysis confirmed that the amplicons detected by the P3 system were the predicted 195-bp satellite repeat specific to the Diaz1/Diaz2 primer set[69], and that the PCR performance is comparable to that of a laboratory thermal cycler instrument (**Figure 2.5b**). These data suggest that our P3 system

should be capable of detecting clinically relevant concentrations of *T. cruzi* gDNA directly from the blood of patients infected with Chagas disease.

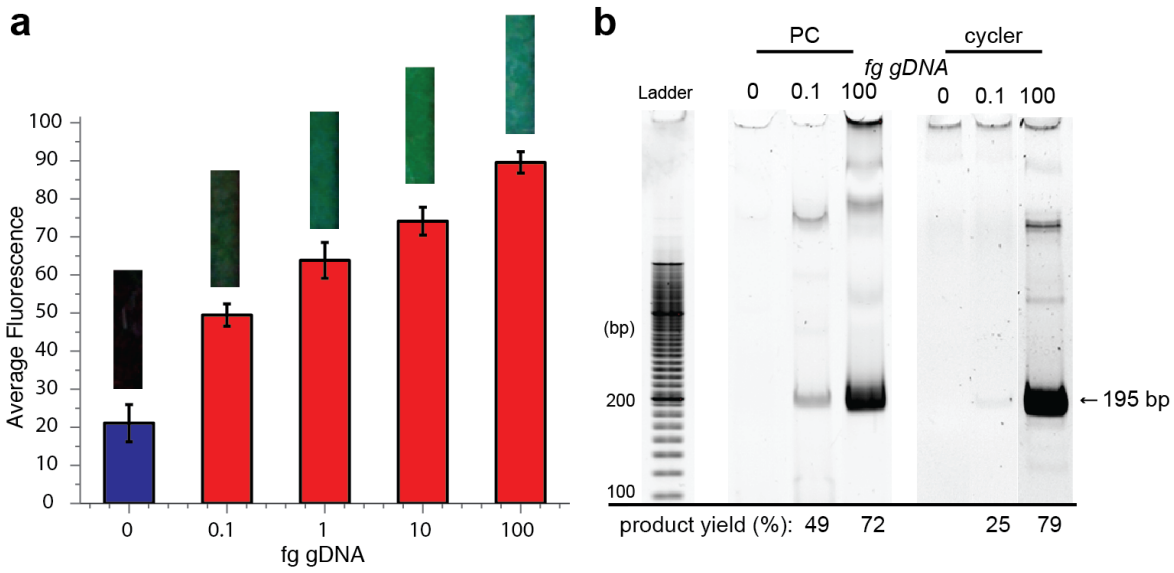

Figure 2.5 Detection of *T. cruzi* gDNA in whole blood. **a.** Images of capillary tubing positioned above bar graphs indicating the fluorescence signal obtained from each amount of template gDNA after background subtraction. P3 is sufficiently sensitive to reproducibly detect 0.1 fg of gDNA in 1 μL whole blood. Error bars were obtained from 4 replicates. **b.** Gel analysis shows the specific amplification of a 195-nt region of tandemly repeating gDNA. A side-by-side comparison of the same sample after PCR carried out in a PC or a laboratory thermal cycler shows that PC-PCR can match the efficiency of laboratory instrumentation.

Conclusion

In this work, we report the feasibility of transforming a desktop computer and cell phone into a molecular diagnostic system capable of highly sensitive and quantitative pathogen DNA detection. These devices are becoming increasingly available in developing countries, especially with the aid of humanitarian organizations dedicated to expanding the reach of technology in areas of need[70,71]. Using two programs, one free and the other affordably priced, we are able to convert a PC into a highly responsive thermal cycler; by adding a simple filter, we are likewise able to use a standard camera phone for the quantitative optical detection of PCR amplicons. We showed that the P3 system is capable of achieving

sensitivities and specificities comparable to that of conventional laboratory instruments at 1/500th of the cost (Table 2.1). In an initial demonstration with the Chagas disease pathogen *T. cruzi*, we achieved a limit of detection for gDNA that is four-fold below the average clinical concentration typically found in patients. There are a number of advantageous features inherent to the P3 system. First, sample prep and handling is minimal, as our assay is performed directly in blood with no need for mechanical or chemical processing and only a brief boiling step required as a prelude to PCR. Additionally, P3 greatly reduces the potential for cross-contamination through the use of a single, sealed capillary tube. Finally, our system allows multiple samples to be analyzed at once (up to 29 in the current configuration), and we envision that it can be adopted for multiplexed detection of multiple pathogens in separate capillaries.

	Fixed Costs		
	PCR	P3	Vendor
Computer	-	$-	
Cameraphone	-	$-	
Blue LEDs	-	$8.00	DX
Light filters	-	$33.50	Edmund
Thermal cycler	$7,200.00	-	Eppendorf
Electrophoresis system	$451.00	-	Bio-Rad
Power supply	$350.00	-	Bio-Rad
Gel documentation	$10,861.19	-	Joyfay
Total	$18,862.19	$41.50	

	Marginal Costs		
	P3	10 µl rxn	
	Purchase price	**Cost/rxn**	**Vendor**
Polymerase	$90.00	$0.450	New England Biolabs
Primers	$5.40	$0.004	IDT
Tubing	$8.00	$0.007	

SYBR Green I	$332.00	$0.000	Sigma
Control DNA	$5.45	$0.000	IDT
Pipette tips	$28.86	$0.03	Fisher SureOne
Total	$469.71	$0.49	

Table 2.1 Tally of fixed and marginal costs associated with amplifying and quantifying DNA according to standard protocols vs. the P3 approach, which can be carried out at $0.49/reaction.

The sensor technology presented here represents a rapid and facile means of pathogen detection in low-resource settings, as indicated by the detection of gDNA in blood. In its current iteration, we anticipate that P3 will work primarily as a binary diagnostic assay in which an endpoint sample readout is compared to positive and negative controls to determine the presence of pathogenic infection. Though beyond the scope of this work, we envision that future iterations of system will demonstrate equivalent performance with patient blood samples containing intact parasites. Several examples in the literature have already established precedents for successful PCR-based detection in blood samples without DNA extraction directly from pathogens such as parasites[72], viruses[73], as well as trypanosomes[74]. We are therefore confident that P3 can achieve a similar level of performance with minimal modifications through techniques such as heat pretreatment and optimization of buffers to facilitate the release of gDNA from lysed pathogens. Finally, we are currently investigating the potential for adapting P3 for use in more remote regions. As resource-limited settings may not have access to reliable power sources for refrigeration or computational operation, lyophilized reagents[49] and battery-powered laptops could be implemented in future iterations. Laptop CPUs have the added advantage of functioning at even higher temperatures than desktop computers with increased portability. Similarly, low-cost blue LEDs and a blacked-out cardboard box can replace UV transilluminators while blocking ambient light for a more inexpensive method of dye excitation (peak absorption

wavelength = 500 nm). In sum, we report an innovative method for transforming widely available consumer electronics into tools for molecular biotechnology. We believe such re-purposing may offer versatile strategies for providing molecular diagnostic capabilities to resource-limited settings in a cost-effective manner for a diverse spectrum of diseases.

Experimental Section

A. Materials and Reagents

All synthetic DNA sequences were purchased from Integrated DNA Technologies (IDT). Hot-start polymerase master mix polymerase and nuclease-free water were purchased from Promega. PCR validation was performed with a 100-nt single-stranded DNA sequence reported previously as Thr-02[24], using primers AGCAGCACAGAGGTCAGATG and TTCACGGTAGCACGCATAGG. DMSO was obtained from the American Type Culture Collection and used at the concentration indicated. SYBR Green I was obtained from Life Technologies and used at 0.625x concentration and EvaGreen was obtained from Biotium and used at 1x concentration. Whole blood preserved in EDTA was obtained from Bioreclamation. Hemo Klentaq (HKT) polymerase and 5x buffer were obtained from New England Biolabs and used in the manufacturer recommended amount. Melting temperatures were measured via iQ5 real-time multicolor detection system (Bio-Rad).

B. PC-based PCR

For initial validation experiments, we used 2 fmol synthetic oligo as template, with 1 μM primer in a 50 μL reaction that included PCR mix (containing DNA polymerase, dNTPs, buffer), EvaGreen, primers, and 13% DMSO. The optimal DMSO concentration was

determined prior to carrying out PCR in a PC by testing the effect of DMSO on dsDNA hybridization in the reaction mixture described above. Melting temperatures were determined by performing 40 cycles of two-step PCR on the 100-nt template to generate double-stranded products in the iQ5. Post-amplification, we carried out a thermal gradient beginning at 65 °C and increasing to 85 °C at a rate of 1 °C per minute and dwell time of 10 seconds, with fluorescence intensity measured every minute. We then calculated the negative first derivative of the plot of the resulting melting profile of intensity versus temperature. The melting temperature is where this differential plot reaches a maximum, as calculated by Bio-Rad iQ5 melt curve analysis. To perform PC-based PCR, the reaction mixture was pre-loaded into short capillaries of perfluoroalkoxyl (PFA) tubing. After adding template, the capillary tubes were permanently sealed at both ends by epoxy (Devcon) to yield contaminant-free testers that are amenable to high-throughput manufacturing. The capillary tube was then inserted between the CPU heat sink fins of a Dell Pentium 4 desktop computer for cycling. The software used for CPU heating was BurnInTest (Passmark), and the fan control and temperature logging were performed by SpeedFan (Almico). The external CPU temperature was cycled from 83 °C to 55 °C for 20 cycles by manual control. External sample temperature logging and monitoring were performed with a digital thermometer (Fluke) and K type thermocouple (Omega). Thermal measurements were performed by recording the solution temperature inside capillaries with 50 μL of distilled water using a thermocouple probe. At the commencement of each cycle, the Maximum CPU temperature test was executed to initiate heating and the CPU fan and case fans were shut down using SpeedFan. When the CPU temperature reached 83 °C, the test was ended and the fans were set to maximum speed to reduce the CPU temperature to 55 °C.

The resulting PCR product was loaded and run on 10% TBE polyacrylamide gels (Bio-Rad) with a 20bp DNA ladder (Bio-Rad) in 4 °C running buffer. After 15 minutes of gel electrophoresis at 300 V, the gels were stained with Gelstar (Lonza) for ten minutes. The stained gels were imaged on a Gel Logic System using UV transillumination and a 535-nm optical filter (Kodak). The positive control PCR was performed in a commercial thermal cycler (Bio-Rad).

Genomic DNA from *T. cruzi* (Tulahuen strain) was obtained from ATCC. The sequence for the Diaz primer set was CGCAAACAGATATTGACAGA and TGTTCACACACTGGACACCAA[75], which target the 195-bp repetitive element in *T. cruzi* nuclear DNA. Capillaries were prepped with 20 µL of reaction mixture containing HK Taq polymerase, 0.2 mM dNTPs, 1x buffer, SYBR Green I, 0.2 uM primers, and 13% DMSO. After 1 µL human blood was loaded into the capillaries, the capillary tubes were sealed on both ends and heated in a boiling water bath for 2 minutes to simulate cell lysis and gDNA denaturation steps prior to PC-PCR amplification. From an initial annealing/extension temperature of 64.5 °C, the annealing/extension step was reduced by a difference of 3 °C every three cycles for the initial step-down cycling phase. After 15 cycles, we maintained the annealing temperature at 49.5 °C for 40 cycles to complete the amplification phase of SD-PCR (**Figure 2.6**).

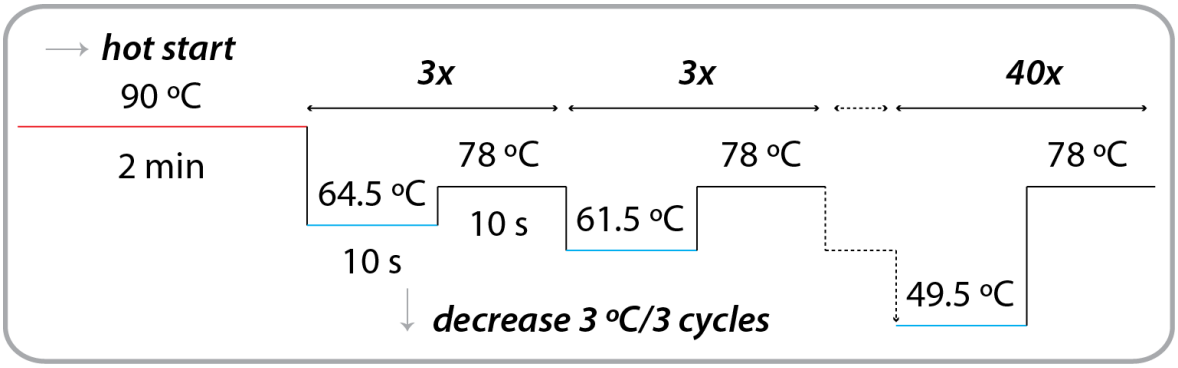

Figure 2.6 Two-step SD-PCR protocol for PC-PCR on *T. cruzi* gDNA. Initial phase is characterized by a stepwise decrease of the annealing temperature by 3 °C increments. The second phase of the thermal cycling profile operates as a traditional PCR amplification.

C. Post-PCR imaging

Samples were excited by a UV transilluminator (Kodak). For imaging, a 520±10 nm bandpass filter (Edmund Optics) was placed over the mobile phone camera (Samsung) and held in place by a silicone case. The phone was situated at fixed distance above the samples and the image was captured using the "night mode" option on the Galaxy S camera phone. Images were transferred to a computer and analyzed with ImageJ software (nih.gov). Rectangular regions of interest were drawn around each sample and the histogram of pixel intensities was obtained. Mean histogram values of all samples were then background-subtracted with the mean histogram value of an empty tube. Average and standard deviations are the result of at least four individual trial runs. Data were imported into MATLAB and plotted. For relative image analysis, the imsubtract function in MATLAB was used to subtract each element in a sample image by the same element in the blank (empty) image. From the resulting RGB values, background-subtracted images were graphed using the imshow function.

References

[1] Y. W. Tang, G. W. Procop, D. H. Persing, *Clin. Chem.* **1997**, *43*, 2021–38.

[2] S. Yang, R. E. Rothman, *Lancet* **2004**, *4*, 337–348.

[3] B. E. Fu, P. Yager, P. N. Floriano, N. Christodoulides, J. T. Mcdevitt, **2011**.

[4] D. Mabey, R. W. Peeling, A. Ustianowski, M. D. Perkins, *Nat. Rev. Microbiol.* **2004**, *2*, 231–40.

[5] C. D. Mathers, M. Ezzati, A. D. Lopez, *PLoS Negl. Trop. Dis.* **2007**, *1*, e114.

[6] P. Yager, G. J. Domingo, J. Gerdes, *Annu. Rev. Biomed. Eng.* **2008**, *10*, 107–44.

[7] B. A. N. Ki-moon, *INFORMATION ECONOMY REPORT 2012*, **2012**.

[8] U. Access, L. C. Technologies, S. Skills, E. Telecommunications, *Int. Telecommun. Union* **2010**.

[9] R. D. Stedtfeld, D. M. Tourlousse, G. Seyrig, T. M. Stedtfeld, M. Kronlein, S. Price, F. Ahmad, E. Gulari, J. M. Tiedje, S. a Hashsham, *Lab Chip* **2012**, *12*, 1454–62.

[10] D. Lee, W. P. Chou, S. H. Yeh, P. J. Chen, P. H. Chen, *Biosens. Bioelectron.* **2011**, *26*, 4349–54.

[11] A. Greenbaum, W. Luo, T.-W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, A. Ozcan, *Nat. Methods* **2012**, *9*, 889–95.

[12] J. Balsam, R. Rasooly, H. A. Bruck, A. Rasooly, *Biosens. Bioelectron.* **2013**, *51C*, 1–7.

[13] D. Tseng, O. Mudanyali, C. Oztoprak, S. O. Isikman, I. Sencan, O. Yaglidere, A. Ozcan, *Lab Chip* **2010**, *10*, 1787–92.

[14] H. Zhu, S. Mavandadi, A. F. Coskun, O. Yaglidere, A. Ozcan, *Anal. Chem.* **2011**, *83*, 6641–7.

[15] H. Zhu, O. Yaglidere, T.-W. Su, D. Tseng, A. Ozcan, *Lab Chip* **2011**, *11*, 315–22.

[16] D. N. Breslauer, R. N. Maamari, N. a Switz, W. a Lam, D. a Fletcher, *PLoS One* **2009**, *4*, e6320.

[17] H. Xie, J. Mire, Y. Kong, M. Chang, H. A. Hassounah, C. N. Thornton, J. C. Sacchettini, J. D. Cirillo, J. Rao, **2012**, *4*, DOI 10.1038/NCHEM.1435.

[18] Y. M. Gomes, V. M. B. Lorena, A. O. Luquetti, *Mem. Inst. Oswaldo Cruz* **2009**, *104 Suppl* , 115–21.

[19] A. Niemz, T. M. Ferguson, D. S. Boyle, *Trends Biotechnol.* **2011**, *29*, 240–50.

[20] J. D. Carolina, *Air Water Borne Dis.* **2013**, *02*, 1–10.

[21] D. P. Manage, J. Lauzon, A. Atrazhev, X. Pang, L. M. Pilarski, *Lab Chip* **2013**, *13*, 4011–4.

[22] R. H. Liu, J. Yang, R. Lenigk, J. Bonanno, P. Grodzinski, *Anal. Chem.* **2004**, *76*, 1824–31.

[23] C. D. Chin, T. Laksanasopin, Y. K. Cheung, D. Steinmiller, V. Linder, H. Parsa, J. Wang, H. Moore, R. Rouse, G. Umviligihozo, et al., *Nat. Med.* **2011**, *17*, 1015–9.

[24] A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. Thomas, H. Sindi, G. M. Whitesides, *Anal. Chem.* **2008**, *80*, 3699–707.

[25] A. M. Foudeh, T. Fatanat Didar, T. Veres, M. Tabrizian, *Lab Chip* **2012**, *12*, 3249–66.

[26] C. J. Easley, J. M. Karlinsey, J. M. Bienvenue, L. a Legendre, M. G. Roper, S. H. Feldman, M. a Hughes, E. L. Hewlett, T. J. Merkel, J. P. Ferrance, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 19272–7.

[27] J. a Lounsbury, A. Karlsson, D. C. Miranian, S. M. Cronk, D. a Nelson, J. Li, D. M. Haverstick, P. Kinnon, D. J. Saul, J. P. Landers, *Lab Chip* **2013**, *13*, 1384–93.

[28] B. S. Ferguson, S. F. Buchsbaum, T.-T. Wu, K. Hsieh, Y. Xiao, R. Sun, H. T. Soh, *J. Am. Chem. Soc.* **2011**, *133*, 9129–35.

[29] A. Rassi, J. A. Marin-Neto, *Lancet* **2010**, *375*, 1388–1402.

[30] P. J. Hotez, E. Dumonteil, L. Woc-Colburn, J. a Serpa, S. Bezek, M. S. Edwards, C. J. Hallmark, L. W. Musselwhite, B. J. Flink, M. E. Bottazzi, *PLoS Negl. Trop. Dis.* **2012**, *6*, e1498.

[31] F. R. Martins-Melo, C. H. Alencar, A. N. Ramos, J. Heukelbach, *PLoS Negl. Trop. Dis.* **2012**, *6*, e1508.

[32] A. Moncayo, A. C. Silveira, *Mem. Inst. Oswaldo Cruz* **2009**, *104 Suppl* , 17–30.

[33] W. A. Al-soud, P. Rådström, **2001**, *39*, DOI 10.1128/JCM.39.2.485.

[34] K. Mullis, F. Faloona, S. Scharf, *Cold Spring Harb.* **1986**, *L1*, 263–273.

[35] K. H. Hecker, K. H. Roux, *Biotechniques* **1996**, *20*, 478–85.

[36] D. J. Korbie, J. S. Mattick, *Nat. Protoc.* **2008**, *3*, 1452–6.

[37] K. Varadaraj, D. M. Skinner, *Gene* **1994**, *140*, 1–5.

[38] N. von Ahsen, C. T. Wittwer, E. Schütz, *Clin. Chem.* **2001**, *47*, 1956–61.

[39] O. C. Moreira, J. D. Ramírez, E. Velázquez, M. F. a D. Melo, C. Lima-Ferreira, F. Guhl, S. Sosa-Estani, J. A. Marin-Neto, C. a Morillo, C. Britto, *Acta Trop.* **2013**, *125*, 23–31.

[40] M. Virreira, F. Torrico, C. Truyens, C. Alonso-Vega, M. Solano, Y. Carlier, M. Svoboda, *Am. J. Trop. Med. Hyg.* **2003**, *68*, 574–82.

[41] Y. Li, N. Kumar, A. Gopalakrishnan, C. Ginocchio, R. Manji, M. Bythrow, B. Lemieux, H. Kong, *J. Mol. Diagn.* **2013**, *15*, 634–41.

[42] Z. Zhang, M. B. Kermekchiev, W. M. Barnes, *J. Mol. Diagn.* **2010**, *12*, 152–61.

[43] R. Ravindran, J. R. Rao, A. K. Mishra, K. Murari, L. Pathak, N. Babu, C. Chencheri, S. Rahul, **2008**, *78*, 89–94.

Chapter III. Quantitative measurement of genetic markers of pathogens in whole blood using a smartphone[2]

Introduction

As the fastest-growing sector within *in vitro* diagnostics, molecular diagnostic technologies are poised to revolutionize the field of healthcare[1–3]. Nucleic acid amplification tests provide exquisite sensitivity and specificity for detecting the DNA and RNA of pathogenic bacteria, viruses, and parasites. However, various challenges have hampered the translation of laboratory-based DNA tests to the point-of-care (POC)[4–6]. Oftentimes, molecular testing relies on bulky, costly, and sophisticated equipment to carry out numerous processing steps (including sample preparation to extract nucleic acids and temperature cycling to amplify the target)[7–9]. Not only are such approaches impractical to perform at the POC, many function as end-point assays that do not provide the quantified copy-number information that is valuable for understanding disease progression and providing accurate diagnosis[10–14].

Numerous innovative strategies have shown particular promise in achieving low-cost, quantitative DNA detection at the POC[15–17]. For example, Stedtfeld et al. developed a miniaturized thermal cycler and phone-based detection system for use in quantifying purified genomic DNA (gDNA) of bacterial pathogens found in environmental[18] and medical testing sites[19]. Jiang et al. also reported an amplification system that was low-cost, solar-powered, and relied on a smartphone for sensing viral DNA[20]. Although these are important advances, they have been unable to predict pathogen load from clinically-relevant

[2] This chapter is in submission for publication as Walker, Faye M.; Fox, Gary N.; Heithoff, Douglas M.; Mahan, Michael J.; Soh, H. Tom. *Quantitative measurement of genetic markers of pathogens in whole blood using a smartphone.*

samples without relying on sample processing procedures that reduce throughput and accuracy[8,21]. Achieving sensitive, quantitative detection of pathogenic DNA at disease-relevant concentrations directly from raw, unprocessed samples remains an elusive goal[22–26].

Here, we report a smartphone-based diagnostic platform ("smaRT-LAMP") that requires minimal resources to provide DNA template copy-numbers from bloodborne pathogens in whole, crudely lysed blood samples. For convenient use at the POC, we developed an Android phone application ("Bacticount") that continuously photographs the DNA amplification reaction, mathematically processes the images, and automatically performs a customized algorithmic analysis that provides the target DNA copy-number without operator intervention. To the best of our knowledge, this is the first study that utilizes the camera feature of smartphones to achieve quantitative DNA analysis in real-time. As a proof-of-concept, we used the smaRT-LAMP system to detect as low as 100 colony-forming units (CFU) of the bacterial pathogen *Salmonella enterica* var. Typhimurium (*S.* Typhimurium) from a single μL of blood in under an hour. We further demonstrated the successful application of our platform for the rapid and sensitive detection of *Salmonella* genomic DNA (gDNA) derived from the whole blood of septic animals[27,28]. The sensitivity of our smart-LAMP system matches that of a laboratory-based (Bio-Rad iQ5) quantitative real-time PCR (qRT-PCR) instrument while substantially reducing the cost and size, and could have a transformative impact in global health applications for limited-resource settings with high disease burden[29].

Results and Discussion

A. Overview of the smaRT-LAMP system

The smaRT-LAMP system is designed to quantitatively detect bacterial pathogens (via real-time optical detection of DNA amplification) directly from blood samples using minimal resources. To start, the blood sample is lysed in a straightforward, single-step process. This crude lysate is combined with a pre-made LAMP reaction mixture that will generate a fluorescent signal in response to template amplification (see **Experimental Section**). Our protocol is convenient, reproducible, and maximizes DNA recovery because it does not require a separate nucleic acid extraction or purification. The prepared LAMP samples are placed in an inexpensive setup consisting of a sample tube holder (that can simultaneously accommodate up to 12 samples), a single-temperature heat block, and an LED light source (**Figure 3.1a**). In comparison to a conventional $35k real-time thermal system, the entire cost for our setup is US$96.

To automatically obtain copy-number information, we implemented our custom Bacticount smartphone application. The Bacticount app was developed by the authors and is freely available through the Google Playstore for phones using the Android operating system (see **Experimental Section**). The operations performed by the Bacticount app are described in **Figure 3.1b**. Briefly, it performs image analysis to recognize the location of each sample tube and obtains its average fluorescence every 10 seconds. With this information, the app employs a "coarse derivative" algorithm to transform the fluorescent data into a time-to-threshold parameter (T_t) that is linearly related to the logarithm of the template DNA copy number[30]. We developed this mathematical algorithm to be quickly completed using a smartphone's processing capabilities and to be robust against fluctuations

in background fluorescence, camera recalibrations, and shifts in the relative position between the sample and the smartphone camera (see **Supplementary Discussion** for further explanation). Finally, the Bacticount app displays the calculated DNA quantities from bacterial pathogens for all 12 samples.

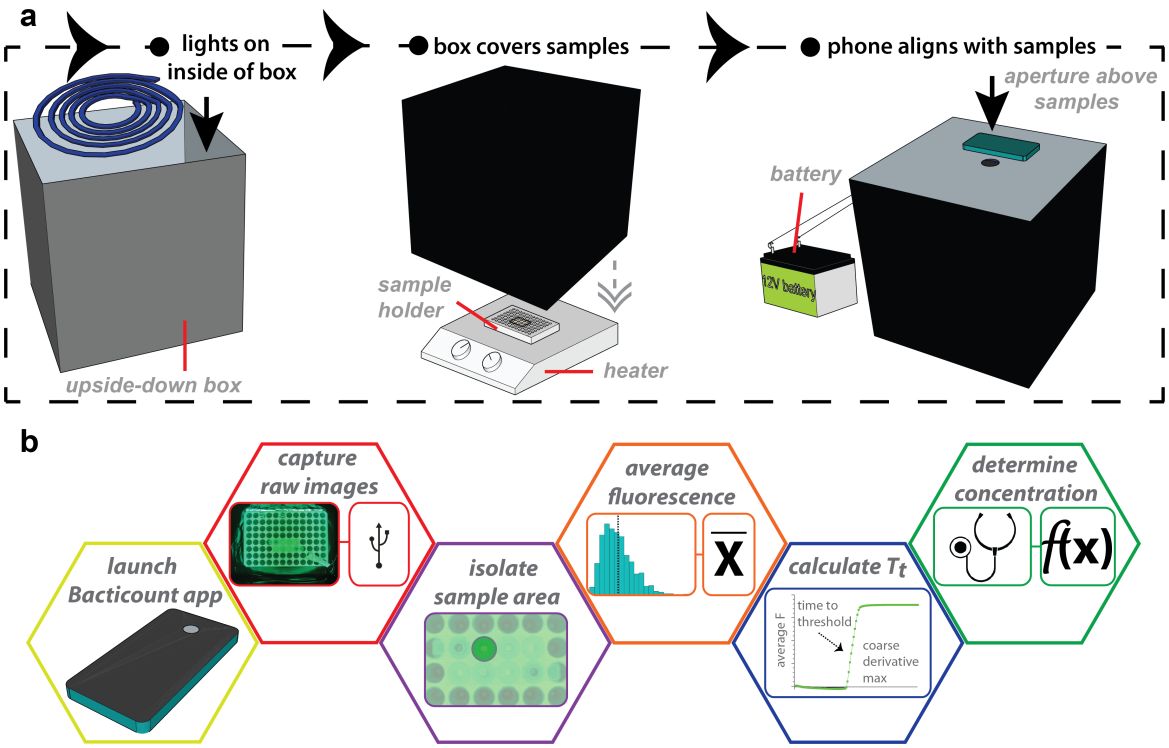

Figure 3.1 The components for smaRT-LAMP detection are a combination of reaction-independent, commonly available materials and reaction-dependent inputs that are determined by the user. **a**. Minimal instrumentation is required, as the platform consists of basic elements. LEDs illuminate the samples, which give off a fluorescent signal that is detected by a smartphone camera outfitted with a green filter. The system is closed off from background light with a cardboard box that is painted black. Once the samples, LEDs, and phone lens have been aligned, the user will be guided through the details of starting a test through a tutorial provided on the Bacticount app. **b**. Diagram demonstrates the algorithm behind the image processing and DNA quantitation carried out by the Bacticount Android app in the course of a smaRT-LAMP run. A photo obtained from the camera function of a smartphone is cropped to isolate an individual sample. A histogram of RGB signals is used to determine the average intensity in the green channel. Raw fluorescent intensities are collected in this manner every 10 seconds for 70-minute run. The data is smoothed before taking the coarse derivative of average fluorescence over the full time period. The maximum of the resulting coarse derivative (roughly analogous to the point of fastest rate of increase) gives the T_t. If the concentration of the sample of interest is known, the T_t will be related to the log of the concentration to form a standard curve; otherwise, the concentration of an unknown test sample can be back-calculated from a reference curve.

B. Accuracy of quantitative profiling

The performance of smaRT-LAMP matches that of a state-of-the-art Bio-Rad iQ5 qRT-PCR instrument in determining precise template copy numbers from pathogenic gDNA. We first determined the detection sensitivity and specificity of smaRT-LAMP by detecting *S.* Typhimurium using a set of six LAMP primers designed to target the highly conserved *recF* gene of *S. enterica*[31,32]. These primers are specific for the model organism, showing no interaction with gDNA from three unrelated parasites in qRT-PCR, melt curve analysis, and qRT-LAMP tests (**Figure 3.2**).

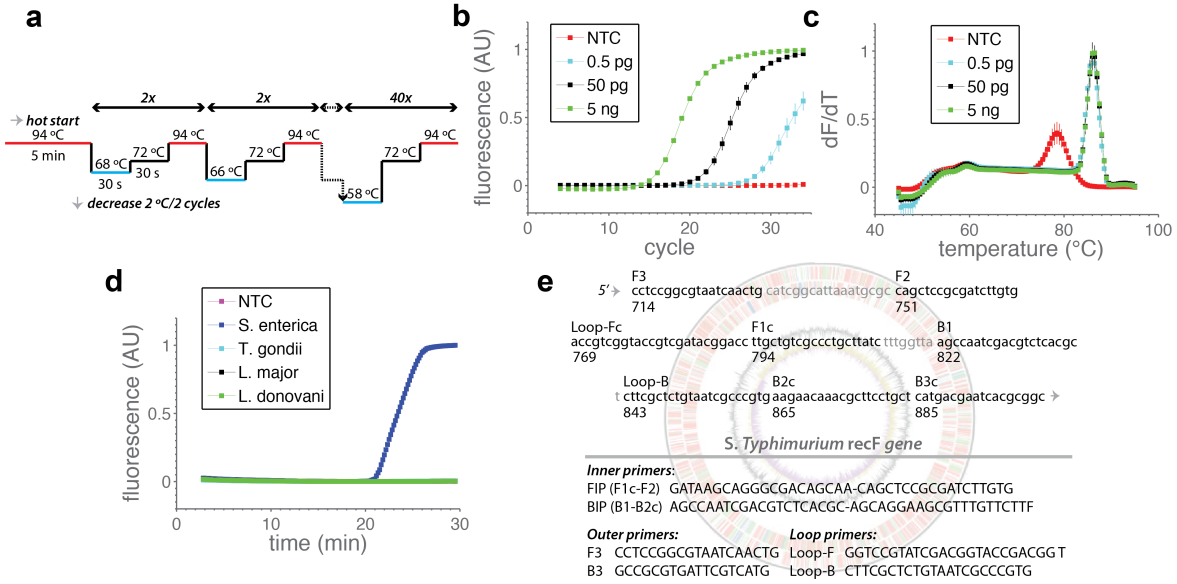

Figure 3.2 Assessment of the sensitivity and specificity of the outer primers, inner primers, and loop primers designed to target the *recF* gene of *Salmonella*. **a**. Schematic of SD-PCR temperature cycling protocol used in amplifying a 190-bp region of the *S.* Typhimurium genome. **b**. Outer primers were defined for use in the LAMP reaction, but are applied to SD-PCR to test specificity for the *recF* gene. Real-time monitoring by EvaGreen shows efficient amplification without spurious by-products. A Bio-Rad qRT-PCR instrument was used to carry out the SD-PCR protocol in **a** on samples with 100-fold dilutions of template gDNA. The limit of sensitivity is 0.5 pg, as illustrated by the amplification quantitation results. **c.** The specificity is such that only one amplicon is generated, as illustrated by the melting curve results. **d**. Performing a LAMP specificity test on a Bio-Rad iQ5 qRT-PCR machine shows that the six primers do not interact with the gDNA of other pathogens. In the presence of three non-target strains of pathogenic genomic DNA, only the desired *Salmonella* species is amplified. **e**. The LAMP primers used to amplify *S.* Typhimurium were adapted from Patterson et al.[31], with the addition of loop primers in the manner of Nagamine et al.[33] F1c indicates the complement sequence of the F1 primer region; similar notation is used for B2c and B3c. The *recF* gene target is conserved in the vast majority of *Salmonella* serovars.

As shown in **Figure 3.3a**, the T_t values and resulting copy-number data from smaRT-LAMP and the standard iQ5 qRT-PCR instrument exhibit minimal differences over a wide range of template gDNA from 100 ng down to 10 pg. The observed T_t values were highly linear with respect to the logarithm of input gDNA in both systems of instrumentation (lines of best fit with regression coefficients both $R^2 > 0.99$ provided in **Figure 3.3b**). Results were accurate and reproducible, as the standard curve from smaRT-LAMP produced average values that were within 4% error of the theoretical concentrations (**Table 3.1**). This strongly demonstrates that the smaRT-LAMP platform could replace a conventional real-time thermal cycler at reduced operative costs.

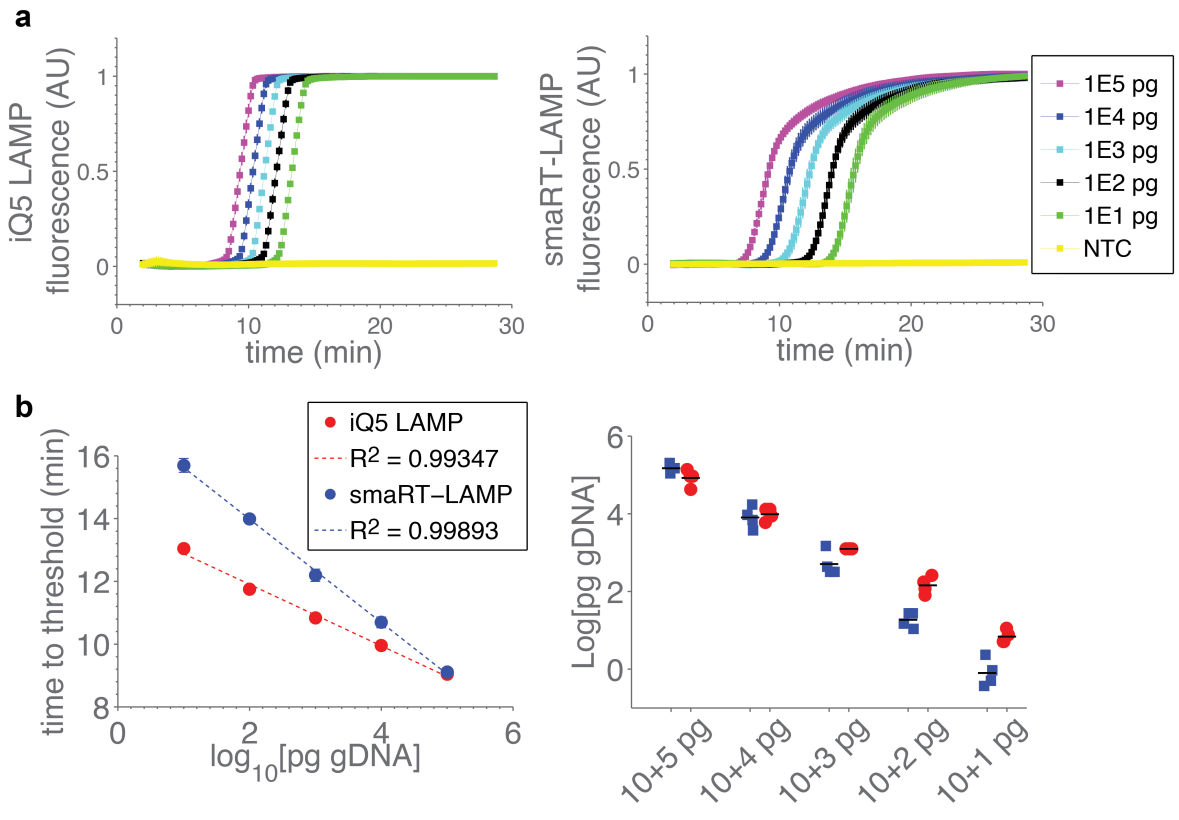

Figure 3.3 Real-time quantification verified with *Salmonella* gDNA. **a**. Real-time signal response from a Bio-Rad iQ5 or smaRT-LAMP system. For the iQ5, fluorescence was measured at 10-s intervals by the cycler, extracted from the optical system software without any additional corrections, and normalized in MATLAB by a custom script. For the smaRT-LAMP platform, the Bacticount app was used to record the green emission of the fluorescent signal every 10 s, calculate normalized

signals, and output a .pasc file of the resulting values. **b.** Signal response of *S*. Typhimurium gDNA with *recF* primers linearly correlates to the log of initial template quantities. Line of best fit from least-squares regression is shown in plot of T_t (red, Bio-Rad iQ5 qRT-PCR system and blue, smaRT-LAMP system) vs. log[gDNA]. The plot of smaRT-LAMP (blue squares) and iQ5 (red circles) gDNA amounts calculated from the best-fit lines show that the measured pg values correlate closely with each other, and with the predicted values indicated on the x-axis. All readings are from 4 determinations with the same sample.

	theoretical log[pg]	5	4	3	2	1
iQ5 qRT-LAMP	experimental Tt	8.83	10	10.83	11.67	13.17
		9.33	9.83	10.83	12	13.17
		9	10.17	10.83	11.5	13
		9	9.83	10.83	11.83	12.83
		0.21	0.16	0	0.22	0.16
	mean experimental Tt	9.04	9.96	10.83	11.75	13.04
	mean experimental log[pg]	4.91	3.98	3.09	2.16	0.85
	% error of mean	1.79	0.48	3.11	8.15	14.77
smaRT-LAMP	experimental Tt	9.28	10.78	12.44	13.78	15.94
		8.94	11.11	12.28	14.28	16.11
		9.11	10.61	12.44	13.78	15.61
		9.11	10.28	11.61	14.11	15.11
		0.14	0.35	0.4	0.25	0.44
	mean experimental Tt	9.11	10.69	12.19	13.99	15.69
	mean experimental log[pg]	4.96	4	3.09	2	0.96
	% error of mean	0.85	0.09	2.87	0.08	3.83

Table 3.1 Initial characterization of smaRT-LAMP technique with purified *S*. Typhimurium gDNA. The log of template quantity was determined by using the coarse derivative algorithm to calculate T_t after performing qRT-LAMP in a standard iQ5 instrument, or in the low-cost smaRT-LAMP setup. For each known concentration of gDNA, four replicate tests were carried out. Calculations assume that the theoretical log[pg] represents the total amount in 1 µL of input sample.

C. Detection of bacterial pathogens in whole blood

To accurately detect nucleic acids from target pathogens, we formulated a procedure that could generate amplifiable gDNA in a crude input sample of whole, lysed blood. Endogenous, blood-based components such as Immunoglobulin G (IgG) or hemoglobin can give off high background signals, quench fluorescent dyes, and inactivate the polymerase in nucleic acid amplification tests[13,34–37]. As LAMP has shown tolerance to inhibitors in crudely processed samples, including case studies with blood[38–45], we bypassed nucleic

acid extraction for a single NaOH[46]/detergent[47]/heat[48] step that would lyse cells by both physical and chemical means. We prepared serial dilutions of cultured ST14028 cells, minimally processed the samples as such, and directly added the lysate to pre-prepared LAMP reagents (**Figure 3.4a**). Fluorescent signals from qRT-LAMP were used to obtain T_t values in the iQ5 and smaRT-LAMP platforms (time-course curves can be found in **Figure 3.5a**). The resulting T_t parameters were linear over several log orders, from 10^2 CFU through 10^7 CFU of ST14028. Furthermore, sensitivities and linearities were indistinguishable between the conventional method and our assay (as shown in **Table 3.2** and **Figure 3.4b** with correlation coefficients of $R^2 > 0.99$ in the least-squares fit for each case).

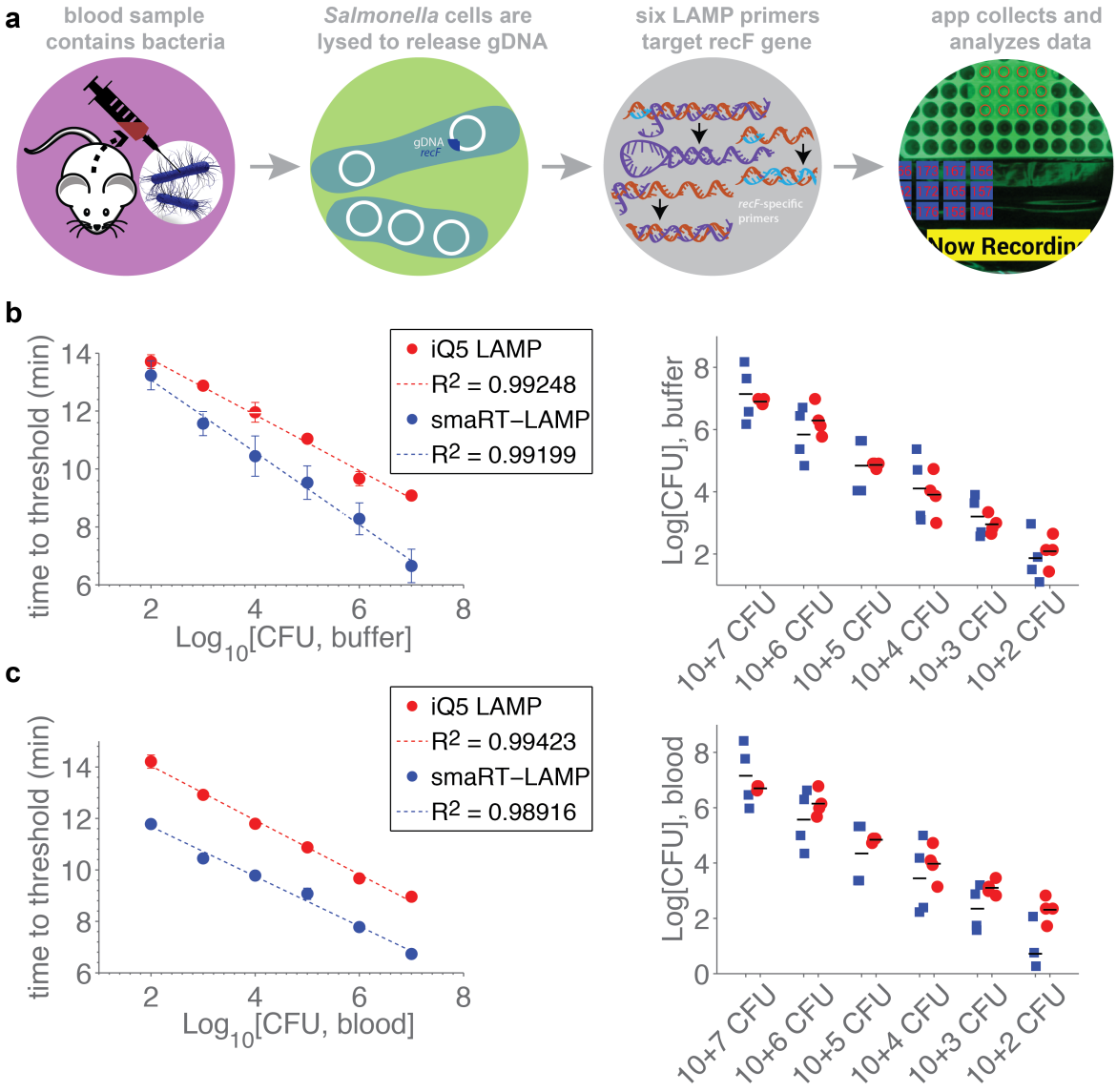

Figure 3.4 Detecting gDNA derived from crude extract of bacteria. **a**. Template samples were lysed with a single-step alkaline and heat lysis, added to LAMP reagents, and amplified at 65 °C. **b**. Amplification of gDNA from ST14028 cells in buffer is carried out in a Bio-Rad iQ5 thermal cycler or the smaRT-LAMP platform. T_ts are calculated by MATLAB in the case of the iQ5 (red), or by the Bacticount app in the case of smaRT-LAMP (blue). Times are plotted against log of known CFU to obtain standard curves with regression coefficients as shown. Measured bacterial counts in buffer are back-calculated from the standard curves, giving numerical values for smaRT-LAMP (blue squares) and qRT-LAMP (red circles) that are approximately equal to the expected values indicated along the x-axis. **c**. Observed same limits and linearity from iQ5 or smaRT-LAMP standard curves after amplification of gDNA derived from bacteria spiked into whole blood. The experimental smaRT-LAMP (blue squares) and qRT-LAMP (red circles) bacterial counts in blood calculated from the T_ts are within range of each other, in addition to being close to the known input concentrations shown along the x-axis. All readings are from 4 determinations with the same sample.

	theoretical log[CFU]	7	6	5	4	3	2
iQ5 qRT-LAMP	experimental Tt	8.67	9.33	11.33	11.33	12.67	13.33
		8.5	9.5	10.67	11.83	12.33	12.5
		8.67	10	10.83	11	12.33	13.33
		8.67	9.33	10.33	11.67	12.67	13.67
		0.08	0.32	0.42	0.37	0.19	0.5
	mean experimental Tt	8.63	9.54	10.79	11.46	12.5	13.21
	mean exp. log[CFU]	7.07	6.09	4.75	4.03	2.91	2.16
	% error of mean	0.96	1.42	5.1	0.79	2.84	7.8
smaRT-LAMP	experimental Tt	9.36	10.03	12.7	12.2	12.53	14.55
		7.49	10.51	10.18	14.2	13.7	15.55
		9.01	9.51	11.03	10.85	16.38	13.7
		8.01	9.01	12.19	11.69	12.19	13.53
		0.87	0.65	1.14	1.42	1.9	0.93
	mean experimental Tt	8.47	9.77	11.52	12.23	13.7	14.33
	mean exp. log[CFU]	7.13	6.07	4.62	4.04	2.83	2.31
	% error of mean	1.92	1.13	7.6	0.93	5.69	15.58

Table 3.2 Limits of detection for smaRT-LAMP and Bio-Rad iQ5 qRT-LAMP in calculating log of initial template quantities from a known concentration of *S*. Typhimurium in buffer. Four replicates were performed for each starting concentration indicated. Calculations assume that the theoretical log[CFU] represents the total amount in 1 µL of input sample.

We next adapted the assay to amplify gDNA derived from known concentrations of ST14028 cells spiked into whole mouse blood. To avoid high background fluorescence from red blood cells, we diluted the input blood template by a factor of 120x into the alkaline buffer used for our NaOH/detergent/heat processing step; the diluted and lysed samples were then used directly for qRT-LAMP in an iQ5 or smaRT-LAMP. Similar results were achieved for the dynamic range in spiked blood as with *Salmonella* bacteria in buffer. Detection sensitivity was 10^2 CFU in either the Bio-Rad iQ5 or smaRT-LAMP platform (**Figure 3.4c**), and no signal was observed from uninfected blood used as a negative control (**Figure 3.5b**). Notably, this limit corresponds to a bacterial load in blood of 10^5 CFU/mL, which is within the relevant range of blood concentrations found in septic mice during late

stage infection[31]. Our results also highlight the accuracy of smaRT-LAMP: average measured bacterial counts, based on the regression curve in **Figure 3.4c**, demonstrate less than 8% error with respect to the theoretical values (**Table 3.3**). This demonstrates the ability of smaRT-LAMP to perform quantification of parasite load in a clinical setting.

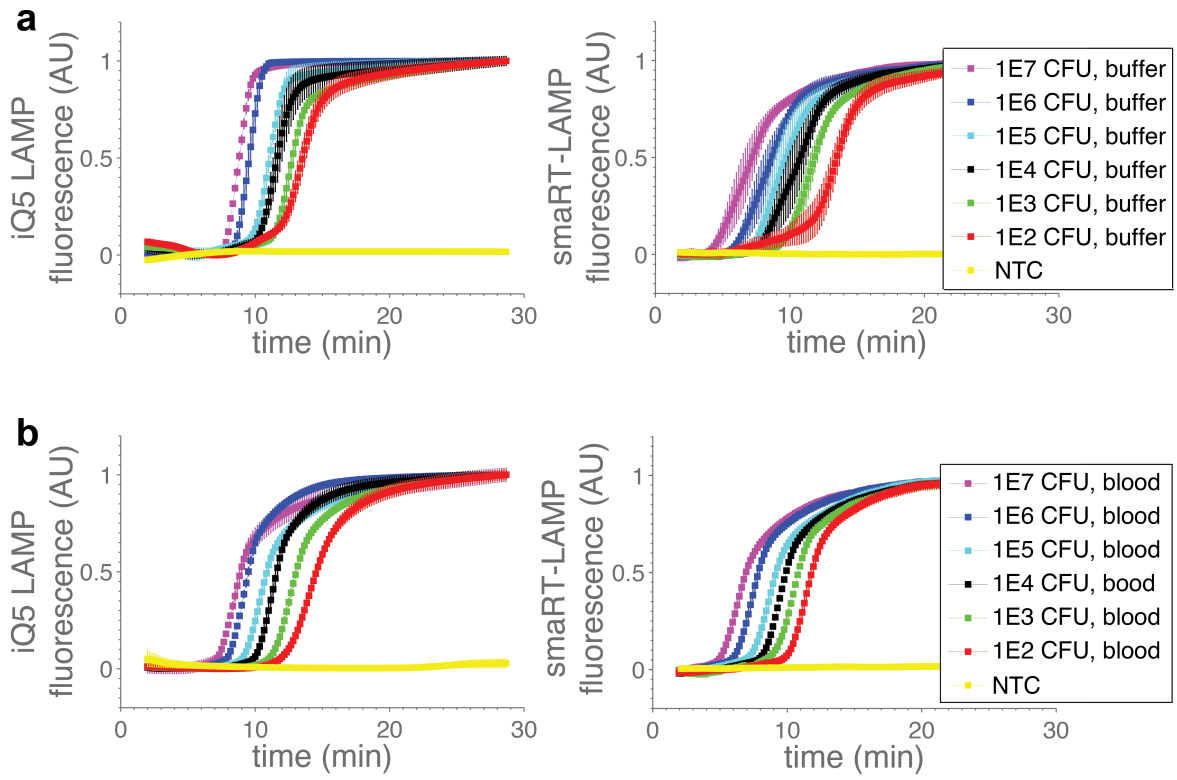

Figure 3.5 Real-time amplification curves obtained from crudely lysed *S*. Typhimurium as template, with error bars from four independent experiments. **a**. Amplification from the *recF* gene of gDNA in crude, lysed *S*. Typhimurium, as detected by a Bio-Rad iQ5 thermal cycler or smaRT-LAMP. Initial template stocks of indicated CFU are diluted, heated, and added to LAMP reagents. Signal from FDR is recorded every 10 s in an iQ5 for real-time plot on left. Raw data is exported and normalized in MATLAB. Smartphone-based smaRT-LAMP platform monitors emission from FDR at 520 nm every 10 s for real-time plot on right. **b**. Amplification of gDNA from *S*. Typhimurium spiked into whole mouse blood using qRT-LAMP or smaRT-LAMP shows clinically relevant levels of detection. Each curve obtained with *S*. Typhimurium indicates the concentration of the template that was initially spiked into whole, uninfected mouse blood. Template samples were lysed, added to LAMP reagents, and monitored at 65 °C in an iQ5 (left) or in smaRT-LAMP (right), and data processed as above.

	theoretical log[CFU]	7	6	5	4	3	2
iQ5 qRT-LAMP	experimental Tt	8.83	9.5	10.5	11.5	12.83	14.67
		8.83	9.67	11.17	11.67	12.83	14.17
		9	9.67	10.83	12	12.83	13.5
		9.17	9.83	11	12	13.17	14.5
		0.16	0.14	0.28	0.25	0.17	0.52
	mean experimental Tt	8.96	9.67	10.88	11.79	12.92	14.21
	mean experimental log[CFU]	6.8	6.14	5	4.13	3.07	1.86
	% error of mean	2.8	2.28	0.05	3.33	2.42	7.21
smaRT-LAMP	experimental Tt	6.94	7.78	9.11	9.78	12.28	11.94
		6.78	7.61	9.44	9.78	11.78	11.44
		6.78	8.11	9.28	9.61	10.11	11.61
		6.44	7.61	8.44	9.94	10.11	12.11
		0.21	0.24	0.44	0.14	1.13	0.3
	mean experimental Tt	6.74	7.78	9.07	9.78	11.07	11.78
	mean experimental log[CFU]	7.06	6.05	4.79	4.1	2.85	2.16
	% error of mean	0.83	0.76	4.2	2.54	5.1	7.89

Table 3.3 Concentrations of *S*. Typhimurium spiked into blood, as assessed by smaRT-LAMP or by a laboratory iQ5 instrument. The log of each initial template quantity was derived via the coarse derivative analysis in four determinations for each indicated concentration of *S*. Typhimurium in blood. Calculations assume that the theoretical log[CFU] represents the total amount in 1 µL of input sample.

D. Quantitative diagnosis of Salmonella *in a murine model of Typhoid fever*

We used the murine typhoid fever pathogen *S*. Typhimurium to evaluate the performance of smaRT-LAMP as a diagnostic platform. Five mice were intraperitoneally infected with a lethal dose (10^3 CFU) of S. Typhimurium parasites. During late stage sepsis (day 5 post-infection), we collected tail blood for smaRT-LAMP analysis. The measured T_t parameters were transformed into numbers of bacteria using the predetermined reference curve of T_t vs. log[CFU] in **Figure 3.4c**. Upon obtaining parasitemia levels via smaRT-LAMP, we performed a comparative analysis with CFU derived from direct colony counting upon plating, as well as qRT-LAMP analysis with the iQ5 (**Figure 3.6**). As shown in **Table**

3.4, the log[CFU/mL] values obtained with our smaRT-LAMP approach were reproducible within replicates and consistent throughout the experiment with those from the Bio-Rad iQ5. In all cases, the two sets of LAMP-based measurements were within 2% difference of each other. Moreover, we established that smaRT-LAMP can differentiate negative from positive cases of infection down to a limit of 2.5 x 10^4 CFU/mL with a rapid turnaround time of 50 minutes per sample run.

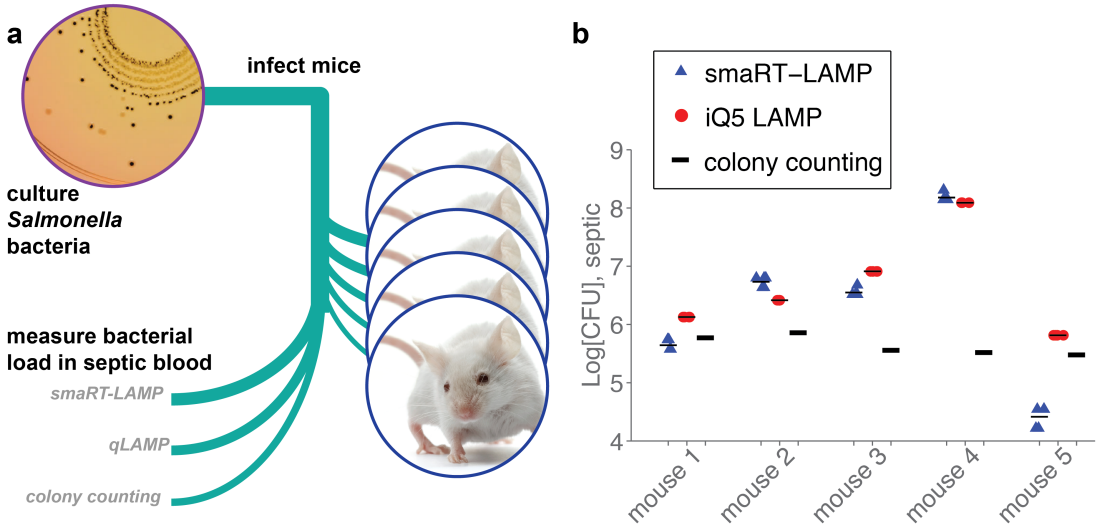

Figure 3.6 Comparison of smaRT-LAMP, iQ5 LAMP, and direct colony counting detection methods from a representative population of *Salmonella* derived from whole blood of septic mice. ST 14028 bacterial cells are from overnight cultures. **a**. A dose of 1,000 CFU is used to infect mice, which reach late stage sepsis at approximately 5 days post-infection. **b**. Mouse blood collected from the tail vein is loaded into the smaRT-LAMP platform (blue triangles), a Bio-Rad iQ5 qRT-PCR (red circles), or reserved for plating and direct colony counting (black bars). Black line represents the mean value from 5 independent runs. Similar limits of sensitivity are observed for all measurements, with predictive values of 1x10^5 - 1x10^6 CFU/mL if biological culture is taken as the gold standard.

	direct colony counting of log[CFU/mL]	5.77	5.86	5.56	5.52	5.48
iQ5 qRT-LAMP	experimental log[CFU/mL]	6.13	6.42	6.91	8.09	5.81
		6.13	6.42	6.91	8.09	5.81
		6.13	6.42	6.91	8.09	5.81
		6.13	6.42	6.91	8.09	5.81
		6.13	6.42	6.91	8.09	5.81
	mean experimental log[CFU/mL]	6.13	6.42	6.91	8.09	5.81
	% error of mean	6.24	9.56	24.28	46.56	6.02
smaRT-LAMP	experimental log[CFU/mL]	5.58	6.64	6.52	8.15	4.55
		5.58	6.64	6.52	8.15	4.55
		5.58	6.8	6.52	8.15	4.22
		5.74	6.8	6.52	8.15	4.22
		5.74	6.8	6.68	8.31	4.55
	mean experimental log[CFU/mL]	5.64	6.74	6.55	8.18	4.42
	% error of mean	2.18	14.95	17.84	48.22	19.38

Table 3.4 The parasitemia level of each mouse at late-stage infection as assessed by smaRT-LAMP or qRT-LAMP in a Bio-Rad iQ5. The log of initial template quantities were computed by converting T_t parameters from a coarse derivative analysis into CFU/mL values via the standard curves obtained from measurements with spiked blood samples. Five determinations were made with the same sample of septic mouse blood for each of the five mice infected with *Salmonella*. Errors of the theoretical log[CFU/mL] results are from comparison with direct colony counts.

Conclusion

In the current study, we have designed a nucleic acid amplification system capable of reliably detecting and quantifying bloodborne pathogens in real time. Our demonstration with non-typhoidal salmonellae, the leading cause of death from food-borne illnesses in the United States[49–51], indicates that smaRT-LAMP is a rapid (signal saturation observed in less than 20 minutes), sensitive (10^2 CFU *S.* Typhimurium per μL of spiked blood), and robust (functions directly on blood specimen) tool for the accurate detection of life-threatening pathogens. The application of smaRT-LAMP to POC monitoring could further global efforts to properly diagnosis salmonella infections and initiate effective therapies,

thus reducing the excessive use of inappropriate antibiotics that leads to untreated morbidity and bacterial resistance[29,52,53]. Our system offers results in minutes, rather than the timeline of 2-5 days required for conventional bacterial culture, and has the added advantage of being a highly automated platform with ultralow costs per assay (less than $2; see **Table 3.5**).

	Fixed Costs		
	iQ5 qRT-LAMP	**smaRT-LAMP**	**Vendor**
Smartphone	-	$-	
Sample holder	-	$65.00	Light Labs
Blue LEDs	-	$6.33	Deal eXtreme
12-V battery	-	$3.97	Home Depot
Green light filter	-	$3.95	Battery Junction
Hot plate	-	$17.00	Aroma Housewares
Real-time thermal cycler	$34,950.00	-	Bio-Rad
Total	$34,950.00	$96.25	

	Marginal Costs for smaRT-LAMP Assays		
	Purchase price	**Cost/rxn**	**Vendor**
Polymerase	$208.00	$0.21	New England Biolabs
Primers	$0.93	$0.01	IDT
Buffer	$52.73	$0.66	In-house
FDR	$128.00	$0.64	SA Scientific
Sample tube	$99.00	$0.10	Bio-Rad
Sample lid	$23.04	$0.02	Bio-Rad
Total for 50 µl rxn		$1.64	

Table 3.5 Comparison of commercial and in-house costs of performing qRT-LAMP. Note that prices can vary according to vendor and purchaser. Tally represents necessary consummables and reagents, but does not account for labor.

With respect to other analytical techniques in the literature, our protocol attains analogous sensitivities for numerical counts of bacterial pathogens. A recent example of a smartphone-based platform for fluorescent microscopy detection of *S.* Typhimurium, developed by Fronczek et al., demonstrated the detection of 10^5 CFU/mL of gDNA spiked

into diluted blood[54]. In real-time measurements of *S*. Typhimurium on an electrochemical device, Patterson et al. demonstrated a sensitivity of 10^3 CFU per sample of infected mouse blood[31]. Our own platform has a lower bound of 10^2 CFU in a single-tube reaction performed on unpurified blood with minimal chemical treatments. In blood samples from infected animals, this corresponded to a lower detection limit of 2.5×10^4 CFU/mL of bacteria. Furthermore, smaRT-LAMP provides amplification test results in real time—with the same reliability, sensitivity, and specificity as a commercially available qRT-PCR system—through the novel application of a standardized "coarse derivative" model of qRT-LAMP.

Our method covers a wide dynamic range with diagnostically relevant limits of detection, correlates well with measurements from laboratory instruments, and has the potential to guide prognostic decisions at the POC. Even with these positive attributes, the smaRT-LAMP system could be further enhanced to maximize its functionality. We anticipate that the high pixel count of cameraphone optics will make it possible to increase throughput without loss of signal resolution. Likewise, we are exploring ways to standardize the setup to encase a heat source, light element, and viewing aperture through 3-D printing or otherwise low-cost fabrication techniques. Looking further to the future, today's availability of whole-genome databases[50,55] will make it possible to adapt smaRT-LAMP to sense a wide variety of other bacterial species and viruses in the detection, profiling, and treatment of bloodstream infections.

Experimental Section

A. Amplification reactions

Previously designed primers targeting the *recF* gene of *S.* Typhimurium[31] were employed with the addition of loop primers chosen to accelerate the reaction by priming strand displacement synthesis[33]. The set of primers consisted of two outer (F3 and B3), two inner (FIP and BIP), and two loop primers (Loop-F and Loop-B). All synthetic oligos were purchased from Integrated DNA Technologies (Coralville, IA). Stepdown (SD)-PCR[56,57] was used to validate F3 and B3 primer sensitivity. Experiments were performed with an iQ5 qRT-PCR instrument (Bio-Rad cat. no. 170-9780) by adding calibration samples containing dilutions of template to a reagent mixture of 500 nM each F3 and B3, 1x EvaGreen (Biotium cat. no. 31000), 1x GoTaq PCR Master Mix (Promega cat. no. M7122), and 5 μM $MgCl_2$ (Sigma-Aldrich cat. no. M1028). Temperature cycling was carried out according to the following protocol in thirty second intervals: from an initial annealing temperature of 68 °C, the temperature of this step was reduced by 2 °C every 2 cycles for the initial stepdown cycling phase. After 10 cycles, we maintained the annealing temperature at 58 °C for 40 cycles of 30 s at 94 °C, 30 s at 58 °C, and 30 s at 72 °C to complete the amplification phase (**Figure 3.2a**). LAMP reactions were carried out in 50-μL volumes containing 0.2 μM each F3 and B3, 1.6 μM each FIP and BIP, 0.8 μM each Loop-F and Loop-B, template (variable), 0.1 μL Triton X-100, 2 μL *Bst* 2.0 polymerase (New England BioLabs cat. no. M0537), 25 μL Loopamp Reaction Mix (SA Scientific cat. no. E-2040; also prepared in-house with no differentiation in outcome[25,58]), 0.8 μL Fluorescent Detection Reagent (FDR, SA Scientific cat. no. E-2210), and nuclease-free distilled water (SA Scientific cat. no. E-2040). All amplifications were carried out at 65 °C for 60 minutes, at which point the temperature was increased to 85 °C for 5 minutes in order to deactivate the polymerase. The smaRT-LAMP platform uses a signal-on approach to quantitatively detect specific DNA sequences.

Measurements of amplicons are based on the enzymatic activity of *Bst* polymerase, which releases pyrophosphate ions as a by-product of DNA synthesis. These pyrophosphate ions complex with divalent manganese ions, releasing the FDR (which contains the fluorescent indicator calcein) from quenching by Mn^{+2}.

B. Preparation of template DNA

Genomic DNA from *S.* Typhimurium strain LT2 was purchased from ATCC (*Salmonella enterica* subsp. *enterica* serovar Typhimurium ATCC-700720). For each 50 µL LAMP reaction, 5 µL of prepared gDNA stock was used as template. Bacterial cultures were prepared from virulent *S.* Typhimurium reference strain ATCC 14028 (CDC 6516-60)[27,28]. For DNA extraction of bacterial samples, a simplified alkaline treatment was used[46]. A 1 µL sample was vortexed for 15 seconds with 49 µL of 50 mM NaOH and 0.5% Triton X-100 and heated at 99 °C for 10 minutes. After cooling on ice, 8 µL of 1 M Tris-HCl (pH 7.5) was added to neutralize. For each 50 µL LAMP reaction, 25 µL of this diluted, neutralized bacteria solution was used as template. In the case of blood samples, steps were performed similarly with the addition of increased sample dilution and vortex mixing. Blood was diluted in a 1:120 ratio with 50 mM NaOH[48] and 0.5% Triton X-100[47]; vortexed for 20 seconds, then an additional 10 seconds; incubated at 99 °C for 10 minutes; snap-cooled on ice; neutralized by addition of 19.2 µL 1 M Tris-HCl; then spun down by centrifugation at 4k RPM for 5 minutes. For each 50 µL LAMP reaction, 25 µL of this diluted, neutralized blood solution was used as template.

C. Salmonella *infection and blood sampling*

Bacterial cells derived from overnight cultures were resuspended in 0.15 M NaCl and delivered to mice via intraperitoneal (i.p.) injection[28]. Whole blood (25 μL) from the tail vein of infected mice was collected on day 5 post-infection into BD Microtainer tubes with K_2-EDTA (Becton-Dickinson, cat. no. 365973).

D. MATLAB analysis algorithm

To develop an image processing pipeline, MATLAB was initially used on the optical measurements from an iQ5 qRT-PCR time-course plot of fluorescence. Background fluorescence is subtracted from the data using the fluorescent values prior to the first inflection point of the curve. Normalization is performed by scaling the fluorescence after the second inflection point of the background-subtracted data so that all values are between 0 and 1. After this basic normalization, we use a processing algorithm (described in detail in the **Supplementary Discussion** section) to find the maximum of the derivative taken over a coarse time stepper. The resultant T_t value is linearly related to the logarithm of the input concentration.

E. Hardware of smaRT-LAMP platform

All experiments were performed in low-profile 0.2-mL PCR strips (Bio-Rad cat. no. TLS-0801) covered with optical flat strips (Bio-Rad cat. no. TLS-0803). For thermal incubation, tubes were placed in an aluminum sample block (LightLabs cat. no. A-7079) on a hot plate. A cardboard box large enough to cover the hot plate was painted black and two flexible cables of 96W, 480 nm, 672 lm, 96-LEDs (DealeXtreme cat. no. 180563) were affixed to the inside top cover of the box. Lights were illuminated with a 12 V battery. A

Motorola Moto X 2nd Generation phone (Motorola Mobility, LLC) was outfitted with a 520±10 nm bandpass filter (Edmund Optics cat. no. 65-699) for visual detection of emitted green light.

F. Development of Bacticount Android application

The Bacticount smartphone application was built on a Motorola Moto X 2nd Generation (Motorola Mobility, LLC) phone using the developer tools in Android Studio (Android). For implementation, the app is downloaded from the Google Play Store and installed via the "OpenCV Manager" application (employed to handle complex algorithms such as image rendering, histogram generation, and back-calculations). Upon opening the app, the user is initially presented with a step-by-step tutorial. Following the tutorial, the user is given a choice to record either a standard curve (reference point calibration) or test reaction. Additionally, the user can view previously processed reactions.

In the case of running a reference or test reaction, the app launches a specialized viewfinder allowing the user to carefully center the reaction vials in the viewframe of the phone's camera such that their intensity can be analyzed over time. Upon hitting "record amplification," the application proceeds to take one photograph of the amplification reaction every 10 seconds over the course of a 70-minute period. The app performs image processing for each of the vials outlined in the viewfinder to extract the average green intensity of each pixel, which is stored in a matrix. For the "standard curve run" option, the software also prompts the user to align each reference sample with ta provided sample map so that the input the starting concentrations of DNA are known. The standard curve is determined through a linear regression fit of T_t vs. log[conc], which is stored as a reference .pasc file for

determining the results in future tests. If the user has selected the "sample run" option, the app will prompt the user to choose a standard curve that has been recorded as outlined in the previous section with known references. After data processing and analysis (described in detail in the **Supplementary Discussion** section), the T_ts of unknown test samples are related to their initial concentrations via the standard curve. On its final screen, the app displays the number of copies of DNA in each reaction vial. The numerical results and collected time-stamped photos are saved as a .parr file and as .jpeg files, respectively, which may be extracted by the user to any computer; alternatively, the digitally-processed results can be uploaded to a central server where a wireless internet connection is available.

All runs are designed with a built-in feature that determines whether amplification has been successful based on the behavior of positive and negative controls. The positive control must fulfill the criteria of having an increase in fluorescent signal that signifies complete amplification, determined in comparison to the negative control. If the positive behaves worse than the negative, the user is notified that the run is a failure. The negative should not amplify at all, and the user is likewise alerted when an increase in amplification is observed.

G. Ethics statement

All animal experimentation was conducted following the National Institutes of Health (NIH) guidelines for housing and care of laboratory animals and performed in accordance with institutional regulations after pertinent review and approval by the Institutional Animal Care and Use Committee at the University of California, Santa Barbara.

Supplementary Discussion

In the LAMP reaction, *Bst* polymerization occurs in a highly parallel and autocyclic fashion that lends the real-time fluorescent output a sigmoidal shape. It is worth noting that methods for describing the kinetic behavior of qRT-LAMP have only started to emerge in the past year[59], and most behavioral models are adopted from the literature of qRT-PCR without being fitted to measured trends[60,61]. For our own purposes, we observed that performing LAMP directly in complex samples, such as bacterial cell lysate and crudely lysed blood, would affect the baseline signal and enzymatic efficiency. Non-idealities from either the smartphone-based detection system or from the endogenous macromolecules in the aforementioned lysates caused noisy fluctuations in the time-course signal. Two smoothing approaches were used to combat this: first, a moving average de-noising filter; second, a coarse derivative technique.

The coarse derivative mechanism is essential in transforming a real-time curve into a T_t value that is linearly related to the logarithm of the input concentration. The first step in manipulating the signal curve is the application of a smoothing filter that averages each point with the ten surrounding point values. Though this will help correct for small changes in the measured fluorescence, it will not account for variations in amplification efficiencies, ground-phase minima, or plateau-phase maxima. To combat such behavioral deviances between samples, we defined T_t as the maximum of the derivative of a coarse model of the real-time curve. Based off of the first derivative maximum (FDM) method[62], the coarse derivative technique avoids false positives that arise from errant noise in the finite first-derivative curve and accounts for sample-dependent variations in amplification efficiencies or time-course minima and maxima.

Instead of computing the signal difference on a point-to-point scale, the coarse derivative takes the difference over a coarse timestepper, *i.e.* a chosen length of time over which to average. The coarse differential of a given system U with time step δt, where time $t_k = k\delta t$, is

$$\frac{\partial U}{\partial t} = \frac{U^{k+1,N} - U^{k,N}}{\delta t}$$

Equation 3.1 Coarse derivative function used to differentiate an output signal U according to a set timestepper δt.

In our case, the optimal δt was empirically determined to be the average rise time from the ground phase to the plateau phase, or 220 seconds. Using the average rise time as a timestepper allows us to compute the timescale derivative of our fluorescence measurements in a way that is more tractable than attempting a finite first derivative, yet commensurate with the measured trends. T_t is hence defined as the maximum signal change over a 220-second time step.

References

[1] R. F. Khabbaz, R. R. Moseley, R. J. Steiner, A. M. Levitt, B. P. Bell, *Lancet* **2014**, *384*, 53–63.

[2] B. (PCR I. Butkus, *Report Values Global MDx Market at $11B by 2015; Projects qPCR to Remain Key Driver*, **2012**.

[3] R. Emmadi, J. B. Boonyaratanakornkit, R. Selvarangan, V. Shyamala, B. L. Zimmer, L. Williams, B. Bryant, T. Schutzbank, M. M. Schoonmaker, J. a Amos Wilson, et al., *J. Mol. Diagn.* **2011**, *13*, 583–604.

[4] P. Craw, W. Balachandran, *Lab Chip* **2012**, *12*, 2469–86.

[5] H. H. Lee, M. a Dineva, Y. L. Chua, A. V Ritchie, I. Ushiro-Lumb, C. a Wisniewski, *J. Infect. Dis.* **2010**, *201 Suppl* , S65–S72.

[6] P. Yager, G. J. Domingo, J. Gerdes, *Annu. Rev. Biomed. Eng.* **2008**, *10*, 107–44.

[7] A. Akane, K. Matsubara, D. Ph, H. Nakamura, S. Takahashi, K. Kimura, R. Akane, **1994**, 362–372.

[8] L. Rossen, P. Norskov, K. Hoimstrom, O. F. Rasmussen, *Int. J. Food Microbiol.* **1992**, *17*, 37–45.

[9] N. Mancini, S. Carletti, N. Ghidoli, P. Cichero, R. Burioni, M. Clementi, *Clin. Microbiol. Rev.* **2010**, *23*, 235–251.

[10] C. Bettegowda, M. Sausen, *Sci. Transl. ...* **2014**, *6*, DOI 10.1126/scitranslmed.3007094.Detection.

[11] H. J. Chung, C. M. Castro, H. Im, H. Lee, R. Weissleder, *Nat. Nanotechnol.* **2013**, *8*, 369–75.

[12] A. Afshari, J. Schrenzel, M. Ieven, S. Harbarth, *Crit. Care* **2012**, *16*, 222.

[13] I. G. Wilson, *Appl. Environ. Microbiol.* **1997**, *63*, 3741–3751.

[14] J. a Platts-Mills, J. Liu, E. R. Houpt, *Mucosal Immunol.* **2013**, *6*, 876–85.

[15] A. Niemz, T. M. Ferguson, D. S. Boyle, *Trends Biotechnol.* **2011**, *29*, 240–50.

[16] P. M. Dark, P. Dean, G. Warhurst, *Crit. Care* **2009**, *13*, 217.

[17] P. Gill, A. Ghaemi, *Nucleosides. Nucleotides Nucleic Acids* **2008**, *27*, 224–43.

[18] R. D. Stedtfeld, M. Stedtfeld, M. Kronlein, G. Seyrig, R. J. Ste, A. M. Cupples, S. A. Hashsham, *Environ. Sci. Technol.* **2014**, *48*, 13855–13863.

[19] R. D. Stedtfeld, D. M. Tourlousse, G. Seyrig, T. M. Stedtfeld, M. Kronlein, S. Price, F. Ahmad, E. Gulari, J. M. Tiedje, S. a Hashsham, *Lab Chip* **2012**, *12*, 1454–62.

[20] L. Jiang, M. Mancuso, Z. Lu, G. Akar, E. Cesarman, D. Erickson, *Sci. Rep.* **2014**, *4*, 4137.

[21] W. A. Al-soud, P. Rådström, *J. Clin. Microbiol.* **2000**, *38*.

[22] T. Cai, G. Lou, J. Yang, D. Xu, Z. Meng, *J. Clin. Virol.* **2008**, *41*, 270–276.

[23] E. T. Han, R. Watanabe, J. Sattabongkot, B. Khuntirat, J. Sirichaisinthop, H. Iriko, L. Jin, S. Takeo, T. Tsuboi, *J. Clin. Microbiol.* **2007**, *45*, 2521–2528.

[24] P. Francois, M. Tangomo, J. Hibbs, E.-J. Bonetti, C. C. Boehme, T. Notomi, M. D. Perkins, J. Schrenzel, *FEMS Immunol. Med. Microbiol.* **2011**, *62*, 41–8.

[25] N. W. Lucchi, A. Demas, J. Narayanan, D. Sumari, A. Kabanywanyi, S. P. Kachur, J. W. Barnwell, V. Udhayakumar, *PLoS One* **2010**, *5*, e13733.

[26] J.-L. Yang, G.-P. Ma, R. Yang, S.-Q. Yang, L.-Z. Fu, a-C. Cheng, M.-S. Wang, S.-H. Zhang, K.-F. Shen, R.-Y. Jia, et al., *J. Appl. Microbiol.* **2010**, *109*, 1715–23.

[27] D. M. Heithoff, W. R. Shimp, J. K. House, Y. Xie, B. C. Weimer, R. L. Sinsheimer, M. J. Mahan, *PLoS Pathog.* **2012**, *8*, DOI 10.1371/journal.ppat.1002647.

[28] D. M. Heithoff, R. L. Sinsheimer, D. a Low, M. J. Mahan, *Science* **1999**, *284*, 967–970.

[29] N. a Feasey, G. Dougan, R. a Kingsley, R. S. Heyderman, M. a Gordon, *Lancet* **2012**, *379*, 2489–99.

[30] Y. Mori, M. Kitao, N. Tomita, T. Notomi, *J. Biochem. Biophys. Methods* **2004**, *59*, 145–157.

[31] A. S. Patterson, D. M. Heithoff, B. S. Ferguson, H. T. Soh, M. J. Mahan, K. W. Plaxco, *Appl. Environ. Microbiol.* **2013**, *79*, 2302–11.

[32] R. a. Edwards, G. J. Olsen, S. R. Maloy, *Trends Microbiol.* **2002**, *10*, 94–99.

[33] K. Nagamine, T. Hase, T. Notomi, *Mol. Cell. Probes* **2002**, *16*, 223–229.

[34] T. Bar, M. Kubista, A. Tichopad, *Nucleic Acids Res.* **2012**, *40*, 1395–406.

[35] B. J. Taylor, K. a Martin, E. Arango, O. M. Agudelo, A. Maestre, S. K. Yanow, *Malar. J.* **2011**, *10*, 244.

[36] Z. Zhang, M. B. Kermekchiev, W. M. Barnes, *J. Mol. diagnostics* **2010**, *12*.

[37] M. B. Kermekchiev, L. I. Kirilova, E. E. Vail, W. M. Barnes, *Nucleic Acids Res.* **2009**, *37*, e40.

[38] K. Hayashida, K. Kajino, L. Hachaambwa, B. Namangala, C. Sugimoto, *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003578.

[39] J. R. Port, C. Nguetse, S. Adukpo, T. P. Velavan, *Malar. J.* **2014**, *13*, 454.

[40] J. Sirichaisinthop, S. Buates, R. Watanabe, E. T. Han, W. Suktawonjaroenpon, S. Krasaesub, S. Takeo, T. Tsuboi, J. Sattabongkot, *Am. J. Trop. Med. Hyg.* **2011**, *85*, 594–596.

[41] R. Howard, J. B. S. Leathart, D. J. French, E. Krishan, H. Kohnke, M. Wadelius, R. Van Schie, T. Verhoef, A. H. Maitland-van der Zee, A. K. Daly, et al., *Clin. Chim. Acta* **2011**, *412*, 2063–2069.

[42] J. H. Chen, F. Lu, C. S. Lim, J. Y. Kim, H. J. Ahn, I. B. Suh, S. Takeo, T. Tsuboi, J. Sattabongkot, E. T. Han, *Acta Trop.* **2010**, *113*, 61–65.

[43] D. H. Paris, S. D. Blacksell, P. N. Newton, N. P. J. Day, *Trans. R. Soc. Trop. Med. Hyg.* **2008**, *102*, 1239–46.

[44] K. a Curtis, D. L. Rudolph, S. M. Owen, *J. Med. Virol.* **2009**, *81*, 966–72.

[45] L. L. M. Poon, B. W. Y. Wong, E. H. T. Ma, K. H. Chan, L. M. C. Chow, W. Abeyewickreme, N. Tangpukdee, K. Y. Yuen, Y. Guan, S. Looareesuwan, et al., *Clin. Chem.* **2006**, *52*, 300–3.

[46] J. Beige, J. Lokies, T. Schaberg, U. Finckh, M. Fischer, H. Mauch, *J. Clin. Microbiol.* **1995**, *33*, 90–95.

[47] D. J. Grab, O. V Nikolskaia, N. Inoue, O. M. M. Thekisoe, L. J. Morrison, W. Gibson, J. S. Dumler, *PLoS Negl. Trop. Dis.* **2011**, *5*, e1249.

[48] M. Soejima, K. Egashira, H. Kawano, A. Kawaguchi, K. Sagawa, Y. Koda, *J. Mol. Diagn.* **2011**, *13*, 334–9.

[49] E. Scallan, R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, P. M. Griffin, *Emerg. Infect. Dis.* **2011**, *17*, 7–15.

[50] S. J. Peacock, *Nature* **2014**, *509*, 557–559.

[51] D. G. Maki, *N. Engl. J. Med.* **2009**, *360*, 949–953.

[52] J. R. Andrews, E. T. Ryan, *Vaccine* **2015**, *33*, C8–C15.

[53] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, et al., *Crit. Care Med.* **2006**, *34*, 1589–1596.

[54] C. F. Fronczek, T. S. Park, D. K. Harshman, A. M. Nicolini, J.-Y. Yoon, *RSC Adv.* **2014**, *4*, 11103.

[55] N. J. Loman, C. Constantinidou, J. Z. M. Chan, M. Halachev, M. Sergeant, C. W. Penn, E. R. Robinson, M. J. Pallen, *Nat. Publ. Gr.* **2012**, *10*, 599–606.

[56] K. H. Hecker, K. H. Roux, *Biotechniques* **1996**, *20*, 478–85.

[57] D. J. Korbie, J. S. Mattick, *Nat. Protoc.* **2008**, *3*, 1452–6.

[58] R. Surabattula, M. P. Vejandla, P. C. Mallepaddi, K. Faulstich, R. Polavarapu, *Exp. Parasitol.* **2013**, *134*, 333–340.

[59] S. Subramanian, R. D. Gomez, *PLoS One* **2014**, *9*, 1–10.

[60] J. Tellinghuisen, A.-N. Spiess, *Anal. Biochem.* **2014**, *449*, 76–82.

[61] A. Tichopad, A. Didier, M. W. Pfaffl, *Mol. Cell. Probes* **2004**, *18*, 45–50.

[62] A. Tichopad, M. Dilger, G. Schwarz, M. W. Pfaffl, *Nucleic Acids Res.* **2003**, *31*, 122e–122.

Chapter IV. Generating aptamer affinity reagents to unlock, isolate, and identify a portion of the Chinese hamster ovary cell proteome[3]

Introduction

For all the progress that biopharmaceuticals have made in creating personalized, efficacious therapeutics, the power of these recombinantly expressed proteins has yet to conquer its greatest adversary: the host cell. Mammalian production cell lines generate recombinant therapeutics to the tune of $113B in annual revenue for the biologics industry[1]. Ironically, the cells also make a complex, heterogeneous, process-dependent population of host cell proteins (or HCPs, which are on the order of 1,200-1,400 for Chinese hamster ovary-derived HCPs[2]) that can cause a toxic immunogenic response in humans. While HCPs may serve important functions within the cells, these reductases, proteases, and glycosylases will degrade, inhibit, or otherwise hinder the functionality of the cell-produced therapeutics[3]. HCPs emerge in the bioreactor either through cell lysis or co-secretion with the product of interest, and must be measured, monitored, and controlled in order to meet regulatory standards by the FDA and other global health authorities. While there are many unique ways of measuring residual HCPs and genomic DNA, it remains an industry-wide challenge to find assays that can suitably and accurately detect HCPs at a level that will meet regulatory requirements.

The test used for monitoring levels of HCP impurities must be sensitive, yet broadly based: these proteins have diverse physicochemical properties and are present at only 1-100 ppm (or nanogram HCPs per milligram of drug substance, ng/mg) in the final stages of

[3] At the date of compiling this thesis, September 2015, this work is still in progress. The contents are thus presented as preliminary results.

downstream processing. Most commonly, an enzyme-linked immunosorbent assay (ELISA) is used to provide numerical results that are amenable for setting control limits and specifications. This format relies on polyclonal antibodies (pAbs)—derived from specially immunized animals treated with a broad array of HCP species—as its HCP detection reagents. Though it is the gold standard in the industry, this approach faces limitations from its sandwich-assay format and its development process. Detection of HCPs requires two high-quality antibodies. The first is immobilized on a surface, where it binds to a single epitope on an HCP; the second recognizes a different epitope and reports a signal. Though this works in theory, measurements are realistically impeded by scenarios in which binding is sterically hindered, only one antibody is generated, or too few binding antibodies are available. Given how profiling HCPs depends on the performance of antibody pairs, it is unsurprising that high-quality preparation of anti-HCP antibodies is critical. Practically, the immunization process is lengthy (9-12 months), expensive, and challenging in terms of HCP coverage[4]. Maximizing antibody coverage also maximizes the procedural difficulty: HCP immunogens are prepared to be free of the therapeutic drug, multiple species of animals (oftentimes rabbits, goats, sheep, or chickens) are immunized, multiple immunizations are applied, and antiserum from different animals are finally pooled after the course of several months. Even with these precautions, it is difficult to obtain full coverage because HCPs that do not cause an immunogenic response in the animal will not generate antibodies and the antibodies that are generated may have low affinity to certain HCPs or be in low abundance. Moreover, a human immune system may react in a more dire magnitude and nature to foreign HCPs than the animals used to generate the pAb reagent[5]. With such caveats

accompanying today's standard operations, achieving better process capability in the biopharmaceutical industry will require a break from conformity and convention.

Manufacturers of biologics recognize that one size does not fit all in HCP evaluation. Orthogonal approaches to traditional antibody-based techniques have been incorporated in analysis and control pipelines, but each specific bioprocess demands a goldilocks approach to finding the right combination of technologies for detecting, identifying, quantifying, and profiling HCPs. In a campaign to capitalize upon the advancements of bioinformatics, protemics, and genomics, we turn to affinity reagents well-suited for the post-genomic age: aptamers. Aptamers are single-stranded nucleic acids with target-binding capabilities. Aptamer generation has become a fast-paced, large-scale endeavor through the integration of genomic tools such as high-throughput sequencing (HTS) and array-based characterization[6]. Yet even as aptamers have been selected for "difficult" protein targets in biological samples[7], the severity of the level of difficulty in HCP detection is marked in the complexity of the Chinese hamster ovary (CHO) cell proteome. Not only are there thousands of constituents in the proteome, the state of each could be anywhere from native to denatured to anything in between[8]. Intracellular processes and extracellular modifications result in charge and size heterogeneity that accordingly confound structural interactions with external binders. Another important factor is the wide dynamic range of the levels of HCPs produced, as protein expression levels in eukaryotic cells can vary over six orders of magnitude[9]. In contrast to the use of traditional *in vitro* selection technology, wherein aptamers are selected for high affinity to one specific target, binding to HCPs requires a novel platform for biomarker discovery. A strategy suited to this task would probe

for multi-target binders that interact with vastly variable content. Selective dominance by a singly enriched family of motifs is no longer the end goal.

In this work, we have leveraged the advent of big data technologies to identify DNA affinity reagents against components of the CHO proteome. This entails successively performing microfluidic selection, particle display screening, HTS, microarray binding assays, and affinity chromatography to create a defined reagent set (**Figure 4.1**). By using nontraditional tools, we are able to achieve maximum diversity in our candidate aptamers and avoid the pitfalls of pool dominance by the most prominent binding motifs[10]. This is clear in our quantitative visualization of the targeted HCPs by two-dimensional difference gel electrophoresis (2D-DIGE): the measured coverage was 70%, a hair's breadth from the pAb (a proprietary batch of CHO HCP antigens harvested by Medimmune, LLC using immunization methods) coverage of 72%. In the interest of characterizing HCPs in a manner amenable to the workflow in recombinant therapeutic protein production, we used our aptamer reagent set to isolate target HCPs through affinity-based chromatography and characterized the selected proteins via label-free liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). The differing physical attributes of bound HCPs—calculated isoelectric points (pIs) between 4.4-9.7, molecular weights (MWs) between 22-328 kD, and proposed functions ranging from proteases to protein binders to lipid binders—lend further credence to the broad range of targets to which aptamers can bind. As aptamer reagents have the advantages over their sera-based counterparts of being more reproducible, tunable, and affordable, we believe they are more amenable to *in vitro* optimization for targeting HCP than pAbs. Our demonstration of the sensitivity and specificity of aptamers in

robustly and reliably detecting HCPs is a first step in moving beyond what is currently possible to produce biologics that are certifiably safe.

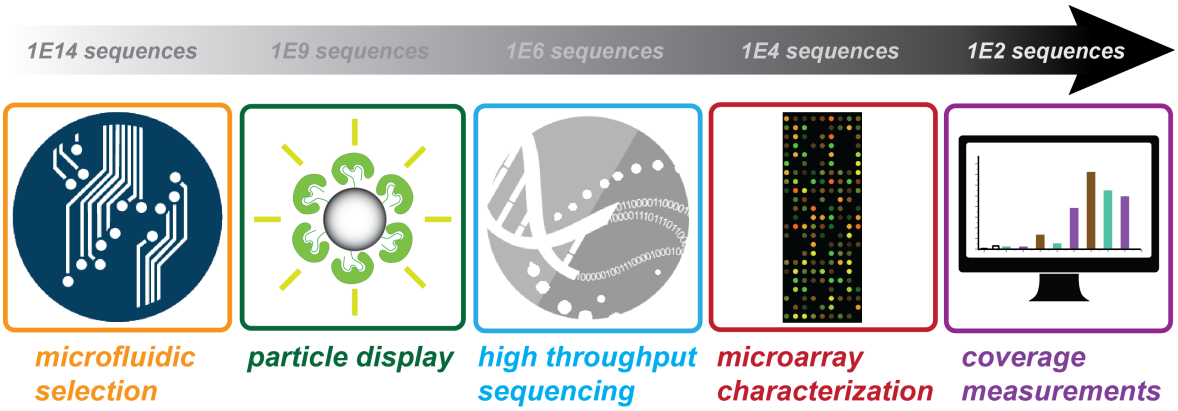

Figure 4.1 Integrative approach for identifying aptamers to HCPs. Workflow is a confluence of technologies for genomics and proteomics. First, a DNA library undergoes directed evolution by microfluidic selection. Next, candidate aptamer pool is screened for target binding by particle display. HTS is used to obtain sequences of selected candidates. DNA microarrays are synthesized and subjected to binding tests with target HCPs to further refine candidate pool. Finally, 2D-DIGE and orthogonal methods are used to test the analytical performance of our aptamers in comparison to pAbs.

Results and Discussion

A. Using confluent technologies to generate candidate aptamers

Our search for a set of effective anti-HCP aptamers used several selection-based screening approaches to obtain a defined reagent set of HCP-binding DNA sequences. The process of systematic evolution of ligands by exponential enrichment (SELEX) selects for target-binding sequences from a library pool with randomized regions by exerting evolutionary pressure that will favor those sequences with high affinity and specificity for the target[11,12]. Most *in vitro* selection platforms are low- or medium-throughput, being limited by the requirement to individually and serially measure the affinity and specificity of each candidate aptamer to a single, known target. In marked contrast, we modify the powerful quantitative parallel aptamer selection system (QPASS) to identify aptamers *en*

masse—while operating on over 1,000 targets simultaneously. QPASS entails minimally enriching a candidate pool of aptamers, sequencing the entire pool, and performing microarray-based binding assays. The introductory application of QPASS by Cho et al. focused on probing and profiling a single target, generating a large collection of sequencing and binding data. Our own extension of QPASS, in expectation of the vast data sets generated in this framework (millions of base reads in HTS and tens of thousands of fluorescent signals on arrays), fully utilized the system's greatest asset of total, parallelized characterization by profiling an entire proteome. The initial pool of 1 nmol (equivalent to $6x10^{14}$ molecules) was designed with a 50-mer random region and enriched through two rounds of microfluidic SELEX (M-SELEX) against HCPs immobilized on beads[13,14]. The protocol diverged from conventional selection with subsequent rounds of selection that used the particle display (PD) technique of fluorescence-activated cell sorting (FACS) to screen for binding of bead-affixed candidate sequences against free HCPs[15]. PD is capable of isolating high-affinity binders from $>10^8$ unique sequences. Its reliance on FACS screening, a high-throughput method to visualize target binding, allows individual sequences to be discarded or retained in real time.

Five total rounds were completed—rounds 1 and 2 (R1 and R2, respectively) using M-SELEX and rounds 3, 4, and 5 (similarly named R3, R4, and R5) using PD—when we began to examine the population of aptamer candidates. Flow cytometry was used to characterize the binding affinities of the rounds after enrichment via PD (**Figure 4.2a**). R5 showed a dramatic increase in affinity for the target protein mixture, as evidenced by the fluorescence intensity from proteins captured by the aptamers (**Figure 4.2b**). The percentage of the R5 population that binds to HCPs, even at low target concentrations of 3 nM, is larger

than both R4 and R3 and indicative of successful discrimination against non-binders in the selection (**Figure 4.2c**).

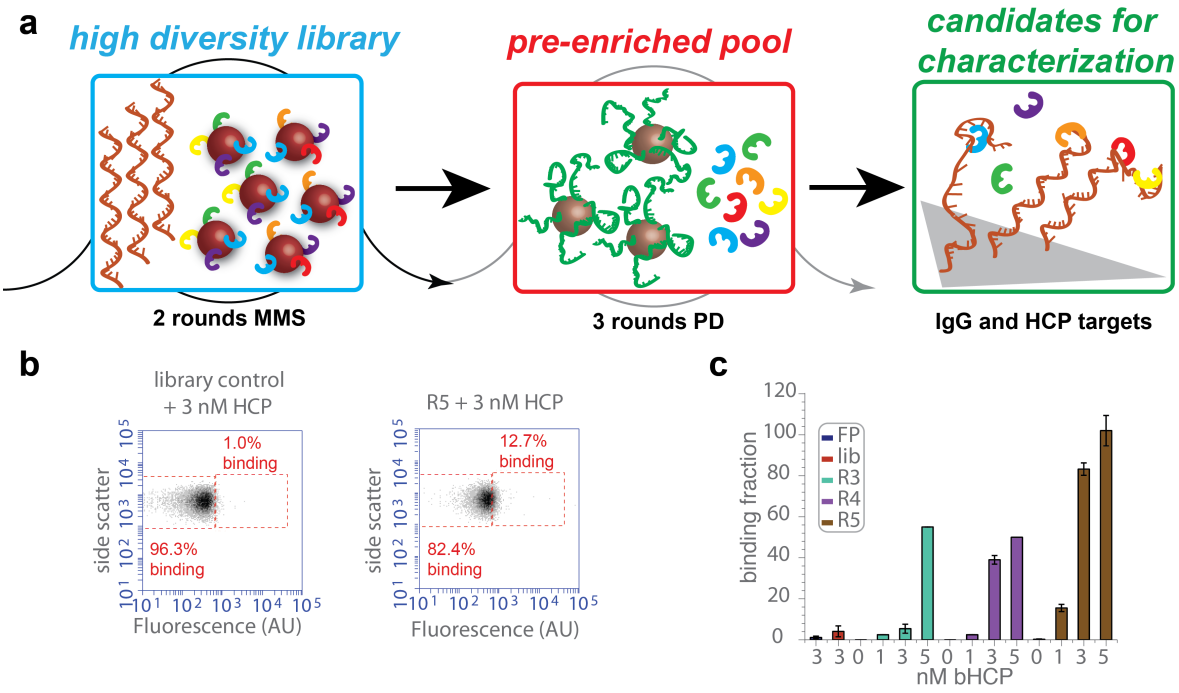

Figure 4.2 Workflow for aptamer enrichment and subsequent tests for target affinity. **a**. After two rounds of SELEX with conventional methodology, we performed further aptamer screening with three rounds of nontraditional PD. **b**. Flow cytometry of enriched aptamer sequences shows reference gates where aptamer particles reside abutting on the sort gate where binding sequences reside. The naïve library control shows only 1% of the population in the sort gate, but 13% of the R5 candidate pool is in the sort gate. **c**. Flow cytometry with 0-5 nM HCP against R3, R4, and R5 pools illustrate that the binding fraction of the population increased accordingly with the level of enrichment in the pool.

Interactions with HCPs represent only one of the qualities that we desired in our aptamers—a preferential specificity for HCPs over other proteins is essential for a reagent set that will need to discriminate against high background levels of biomolecules when applied to process characterization studies. Again relying on flow cytometry, we challenged our R5 pool with increasing concentrations of immunoglobulin G (IgG). IgG is the most common antibody type, recognizing pathogenic viruses, bacteria, and fungi; it is often used

as a standard in internal quality control assays in conjunction with anti-human IgG monoclonal antibodies. If the aptamer pool's performance could stand the test of retaining binding functionality of variable compounds, it would lend further credence to the selective nature of our affinity reagents. IgG from several drug makers (Sigma-Aldrich Co. and EMD Millipore) was introduced at concentrations over 1,000 times higher than HCPs in binding experiments (**Figure 4.3a**). As evidenced by the binding fraction that was retained by HCP, aptamers exhibited affinity for the HCP target in all tests and proved able to retain specificity in scenarios resembling processing characterization. Given this increase in binding ability, we cloned the R5 pool into *Escherichia Coli* (*E. Coli*) and picked 119 clones; of these, 83 were unique. The sequences that were highly represented in families were further used to obtain specificity profiles against IgG using printed microarrays (representing Families I, II, and III in **Figure 4.3c**). Microarray testing was carried out on identical DNA-spotted slides—consisting of an epoxy substrate coupled to amino-modified oligonucleotides—with either HCP or IgG as the target of interest. Binding to HCP was evident from the fluorescence signal observed after scanning for the labeled target. Any signal from IgG was solely due to the intrinsic fluorescence of the substrate and background noise from the solution (**Figure 4.3b**). This survey with a limited number of sequences opened the door to a wider set of characterizations with more sophisticated analysis tools: HTS and high-density arrays.

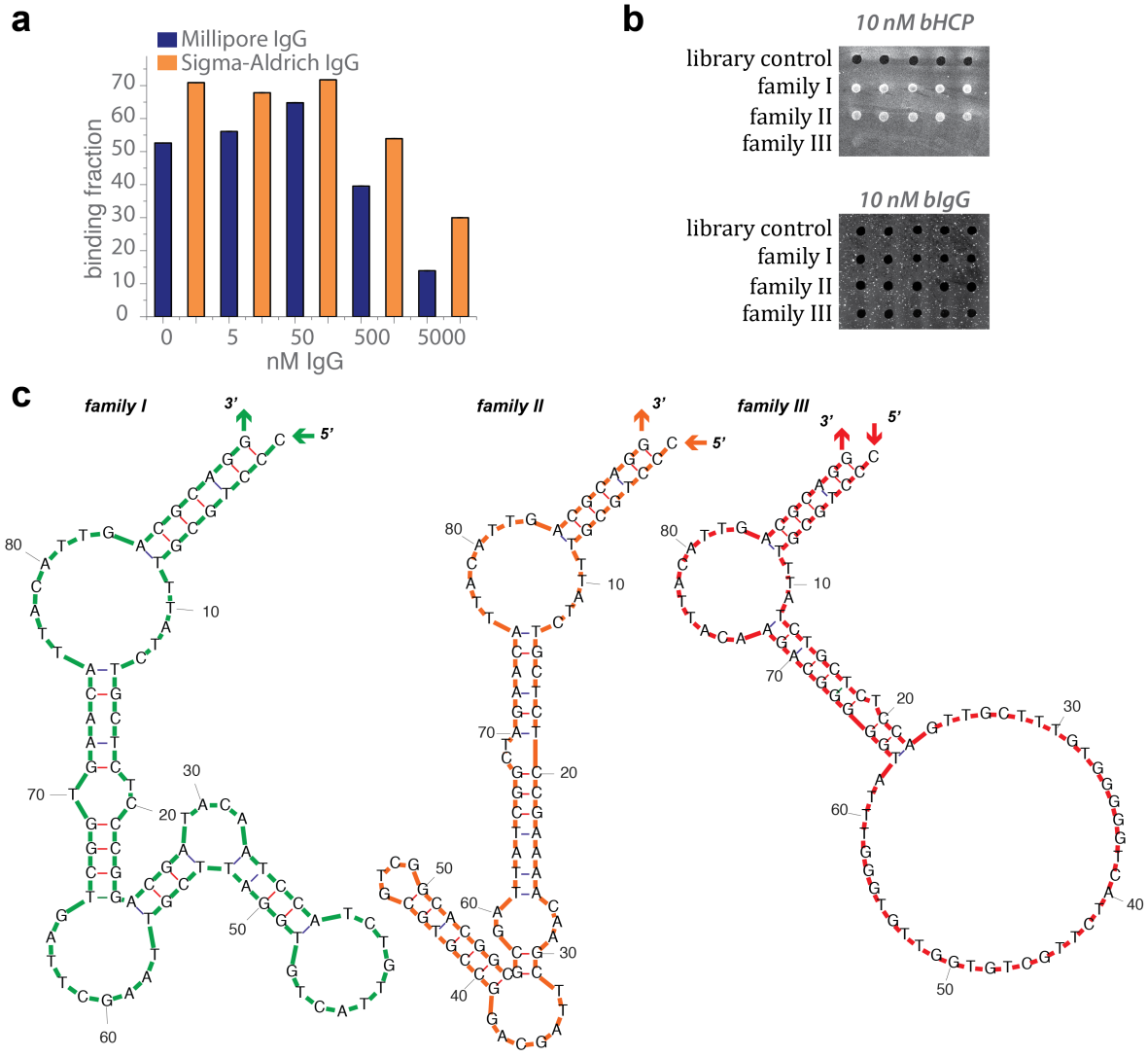

Figure 4.3: Specificity tests against IgG with HCP aptamers from R5. **a**. R5 candidate pool of aptamers were subjected to flow cytometry, with varying amounts of IgG and 3 nM HCP. In up to 5,000 nM IgG, the bead-bound aptamers still exhibited a fluorescent signal from binding to the HCP target. **b**. Sequences, obtained from R5 cloning and sequencing, of candidate aptamers containing C6 amino modifiers at the 5' end were printed on epoxy-modified glass slides to create printed DNA microarrays. Array slides were blocked, incubated with biotinylated HCP or IgG, and labeled with streptavidin-Alexa 647 to simultaneously characterize the affinities and specificities of candidates in parallel. DNA sequences once more illustrate binding greater than the library to HCPs, while giving negative responses equivalent with the library in response to IgG. **c**. Secondary structure motifs from mfold representative of three families in the alignment of 119 sequences from R5 that were used in array binding tests in **b**.

B. Quantitative binding measurements on array to determine final aptamer reagent set

Our goal herein had been to increase the affinity and specificity of the candidate aptamer pool for HCP while separating out the chaff that would bind IgG nonspecifically. Obtaining the maximal coverage of the CHO proteome, from the low-abundance proteins to more concentrated ligands, was our next goal. There are several reasons for retaining the maximum possible diversity in the pool of potential aptamer candidates. Since we are targeting over 1,000 proteins in the mixture of HCPs, it would be undesirable for the pool to converge to only a few sequences of limited structural motifs[10]. To this end, we left behind the limitations of Sanger sequencing; similarly, we backtracked from our use of the most converged round and focused our first pool from PD. Against the convention of Sanger sequencing, we applied the QPASS practice of HTS on our R3 pool. The maximum possible diversity that we can analyze is hence as large as the hundreds of millions in a MiSeq sequencing platform (Illumina). This approach allows us to probe the DNA sequence space for a large number of interactions in parallel, and for rare interactions that would be entirely missed by evaluating a much smaller fraction of sequences.

As HTS is capable of capturing these extremely low copy sequences and successfully identifying high-quality binders[16], we turned to parallelized characterization as a compass to guide us towards these elusive binding agents. According to convention, the number of reads for each unique oligonucleotide in a sequence pool should correspond to enrichment by target binding. The observed trends are, in practice, significantly affected by PCR bias and other artifacts such that a larger copy number does not necessarily indicate a larger binding affinity[17]. Of the three sequences used in the preliminary test with printed microarrays, the sequence from Family I had the lowest ranking by HTS (23,930 of 218,557). However, it gave the highest signal response in binding tests with HCP, as well as

being specific for HCP over IgG (**Figure 4.2c**). These variations between HTS counts and binding performance are best faced head-on with analytical tools. Arrays continued to be a broad-spectrum filter for this work, as in QPASS, but with the caution that aptamer sequences may behave differently on arrays than they would in the final format used to detect HCPs. Much like the biopharmaceutical industry itself, we relied on orthogonal methods of bead-based pull-downs and affinity chromatography to battle these biases and limitations in selecting the defined set of aptamers to be used in the 2D-DIGE determination of HCP coverage.

Parallel mass characterization of aptamers on microarrays has previously been used to identify potential aptamers from HTS data, and the high-throughput nature of this methodology is well-suited to assays with target proteins that number in the thousands[6]. With the HTS results as a starting point, we performed HCP binding assays on *in situ*-synthesized microarrays containing the top 6,000 sequences by copy number. As our aptamer library contained a 50-mer random region flanked by 20-mer primer-binding sites, we excised the core 50-mer portion of the sequence for *in situ* synthesis on the microarrays. Fluorescent intensities from bound HCPs—as noted in the Family I sequence of rank 23,930—did not decline as a result of decreasing copy number, but consistently demonstrated values above a threshold minimum of 1.5×10^4 (AU) down to the lowest-ranked sequences. For several dozen R5 sequences that displayed a binding signal above this set intensity threshold, we performed further investigations via bead-based (pull-downs) and resin-based (HPLC) affinity measurements. From the pull-downs of capturing HCPs in solution on bead-bound aptamers, six distinct families (Families I, II, and III being the same as above) of 22 sequences performed robustly enough to progress towards a reliable

calculation of HCP coverage (**Figure 4.4a**). Representative sequences from each family show very little internal folding, as secondary structures are limited to stem-loops with, at most, 7 base pairs (**Figure 4.4b**). A lack of prescribed self-folding may make the randomized region more available for intermolecular hydrogen binding with the target molecules. This possibility was put to the test in a gel electrophoresis approach to obtain the extent of HCPs bound by these aptamers.

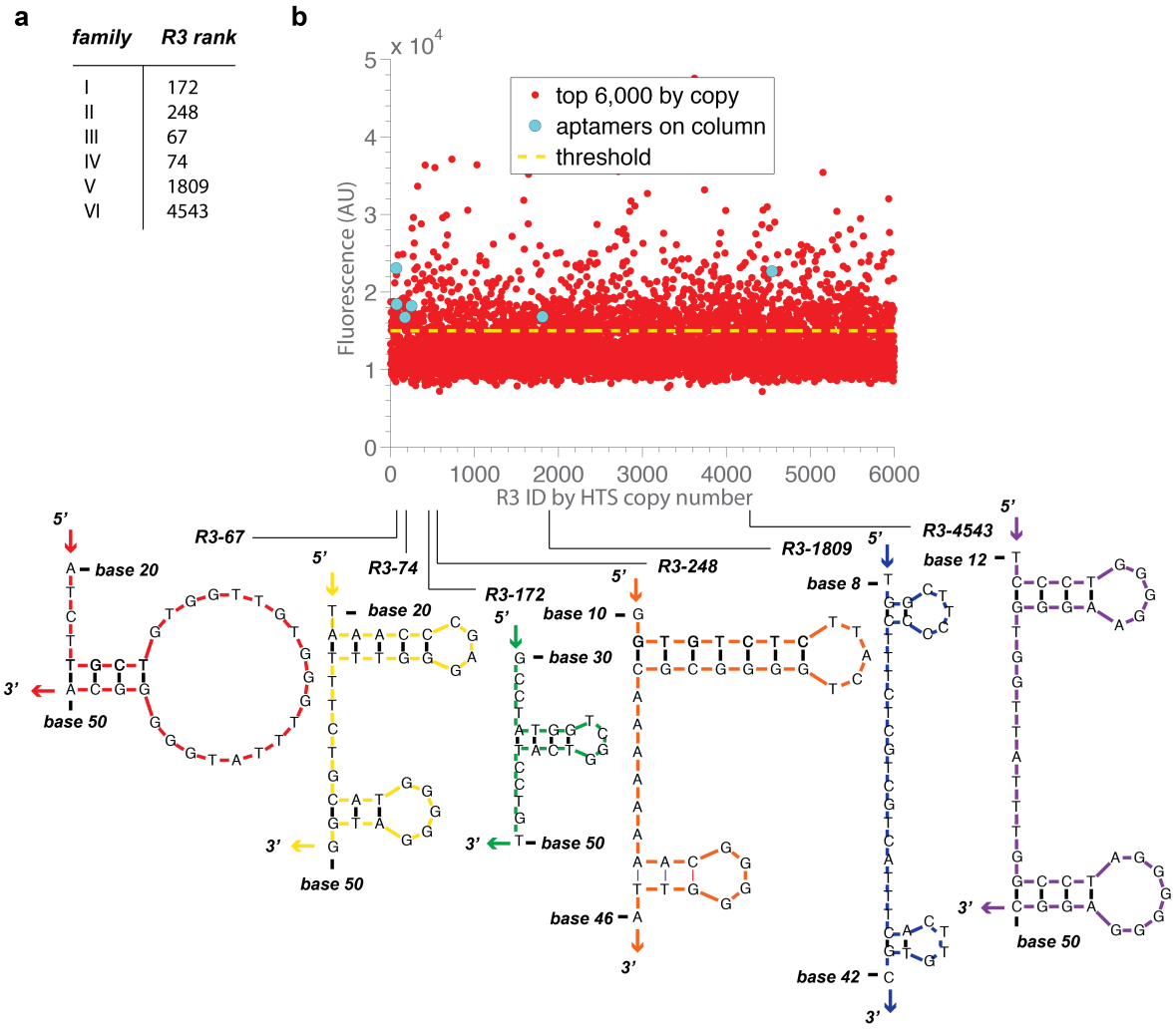

Figure 4.4: The top 6,000 candidate aptamer sequences identified by HTS (according to copy number) were synthesized as 50-mer microarrays to characterize their binding to the HCP target mixture. HTS was performed by GeneWiz. **a**. Aptamer sequences (R3 copy number shown) in **b** that performed well in microarray binding experiments and the orthogonal method of pull-down assays

fell into six distinct families. **b**. After blocking, synthesized microarrays of 50-mer variable regions were incubated with 100 nM biotinylated HCP, washed, and scanned. Fluorescence indicates level of binding affinity. Cutoff for further consideration in affinity-based assays was a signal of 15k. Consensus motifs from mfold for six families as shown (no internal binding is observed in regions of sequences that are not displayed). Microarray synthesis was performed by MYcroarray.

Just as copy number is not predicted to be an accurate gauge of relative binding affinity, the microarray fluorescence measurements did not directly correlate to aptamers with high responses in pull-downs. Two of the R3 aptamer candidates from our preliminary microarray spotting test, representing family II (rank 289) and family III (rank 67), had shown strong fluorescent responses as full-length binders on the arrays but gave marginal signals on *in situ*–synthesized 50mer microarrays. In the face of this discrepancy, we further investigated the capacity in of each in pulling down HCPs as bead-immobilized reagents. Strong, distinct band patterns were observed from HCPs of various shapes and sizes by SDS-PAGE visualization in the case of each pull-down (**Figure 5a**). While it is known that the composition of proteins in a sample of HCPs is widely complex, not much has been accomplished in the interest of deconvoluting said mixture. We completed our progression from library evolution to individual sequences by confirming the identity of bound HCPs via LC-MS/MS. In a comparison with a library control, the two aptamers together bound 11 HCPs (2 by the Family I aptamer and 9 by the Family II aptamer) that had proposed functions ranging from proteases to protein binders to lipid/ion/metal binders as determined by InterPro (http://www.ebi.ac.uk/interpro/) (**Figure 4.5b**). The heterogeneity of the HCPs was likewise evident in the isoelectric points (pIs, between 4.4-9.7) and molecular weights (MWs, between 22-328 kD) (**Figure 4.5c**). It has been previously shown that the pI of a protein has an effect on dissociation constant (K_d)[18]: a protein with a higher pI will be more positively charged, which generally leads to tighter binding against negatively charged

DNA and hence a lower K_d. In contrast to expectations, we observe calculated pIs ranging down to a lower bound of 4.4.

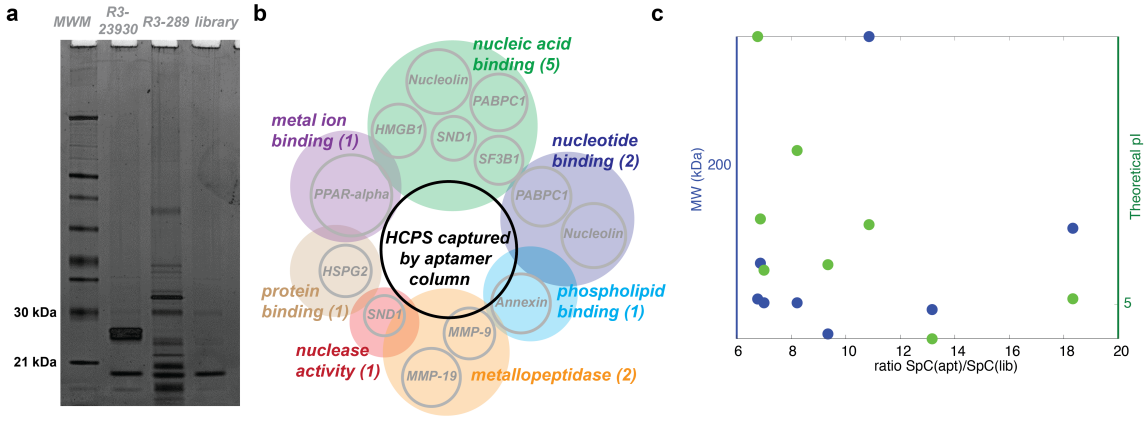

Figure 4.5 LC-MS/MS analysis of DNA-interacting HCPs identified by their form and function. **a**. Aptamer candidates R3-23930 and R3-289 were immobilized on beads, incubated with HCPs, and bound proteins were heat-eluted from the beads. This elution was run on a 4-12% Bis-Tris gel for SDS-PAGE and stained with SYPRO Ruby. **b**. After gel imaging, bands containing proteins were excised and TCA precipitated. Samples were sent to MS Bioworks, where a total of 11 proteins from the R3 aptamers were not observed in a negative control pull-down with the library. Functions shown were obtained via InterPro. **c**. Ratio of aptamer-binding spectral count to library-binding spectral count for statistically significant proteins is shown as a function of pI (right-hand axis) and MW (left-hand axis). Only proteins with a spectral count >50 were included in the analysis. LC-MS/MS was performed by MS Bioworks.

C. Analysis of eluate from aptamer- and antibody-based capture of HCPs

For the application of our aptamers to a proteomic study of CHO HCPs, our platform culminated in 2D gel separation to determine the coverage of the aptamer binders in purifying HCPs from a stock sample. The final set of six aptamer families (**Figure 4.6b**) had been shown to bind HCPs when immobilized as printed or *in situ*-synthesized DNA microarrays, and had passed the criterion of pulling down HCPs in bead-based assays. In determining the extent of HCP coverage, aptamers must be immobilized as a stationary phase in affinity chromatography and actively bind and retain HCPs as they flow over the column bed, from which the HCPs can be quantified by 2-D differential gel electrophoresis

(2D-DIGE) (**Figure 4.6a**). The aptamers best suited for this setting of high-production manufacturing will not necessarily be those that have the largest number of copies present in HTS data or exhibit the highest binding ability on a microarray surface, making our quality check of binding by orthogonal methods a necessity. Total cell lysis was prepared and separated by 2D-DIGE to determine the full collection of HCPs expressed; likewise, HCPs that had been loaded onto an aptamer-conjugated sepharose resin column and eluted after extensive washing were subjected to 2D-DIGE. Many of the spots present on our aptamer-bound sample were not present on a set of library sequences that served as a negative control, indicating the success of our selection process. Moreover, the series of spots bound by the aptamers is similar in magnitude (number of spots) and diversity (range with respect to isoelectric point and molecular weight) to an industry-standardized sheep anti-CHO HCP pAb. A side-by-side comparison demonstrates that the 70% coverage of the aptamer reagent set against HCPs (25 mg load) is nearly equivalent to that of the antibody set at 72% (**Figure 4.6c**).

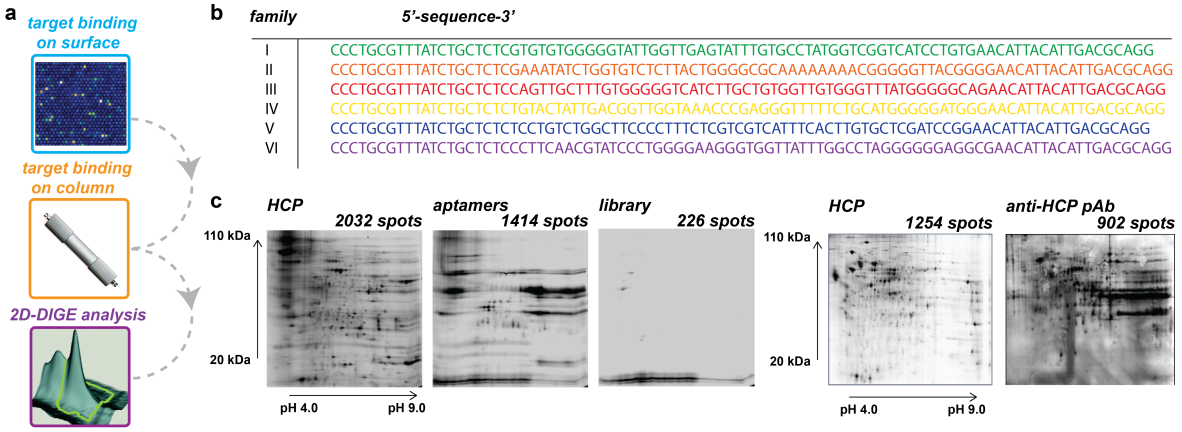

Figure 4.6 Proteomic approach for HCP identification and evaluation to illustrate coverage of aptamers in comparison to pAbs. **a**. Consensus sequences of aptamer families are printed on microarrays to simultaneously characterize the affinities and specificities of candidates in parallel. These results are cross-correlated with aptamer:target binding pull-down assays. **b**. 22 aptamers represented six families of binding motifs, constituting the aptamer reagent set that was used as the

stationary phase of the column in affinity purification. **c**. Amino-modified aptamers or pAbs were coupled to an NHS-sepharose resin. HCP was loaded onto the resin, non-binders were washed away, and bound proteins were eluted under high salt conditions. Fractions were collected and TCA precipitated by MedImmune. 2D-DIGE was performed by MS Bioworks with the naïve library, the 22 selected R3 sequences, and pAbs. Compared to the HCP load, aptamers attained 70% coverage; the pAbs 72%.

Conclusion

By integrating powerful analytical methods, we were able to select DNA aptamers that could in turn be used to extract and identify the recombinant proteins synthesized in CHO cells. Throughout the development process of biotherapeutics, there is an increasingly difficult challenge to quantitate impurity levels that fall to parts per million in the downstream portion of the development pipeline. In testing our aptamers for HCPs against IgG as an industrially relevant drug sample, we demonstrated functional binding to HCPs at levels of the order of 1 ppm. We walked a hard line between maintaining specificity for over 1,200 highly diverse HCPs and avoiding interactions with IgG or otherwise irrelevant biomolecules. Fortunately, our modified QPASS workflow did manage to retain sequences that bound proteins showcasing a varied set of sizes, functions, and pIs: measured molecular masses ran a gamut from 22-328 kDa, and calculated pIs from 4.4-9.7. Having identified these hits for individual HCPs, we further applied a 22-sequence aptamer reagent set to proteomic analysis of HCP by 2D-DIGE separation, wherein the R3 aptamers performed on par with a typical anti-HCP antibody reagent in terms of coverage. In this manner, we can potentially facilitate the predictable and affordable use of nucleic acids to regulate levels of HCPs in cell-produced therapeutic drugs. Our own study was limited to a small set of aptamer reagents hand-picked from over one million unique sequences identified in the R3 pool, indicating that ongoing efforts to perform pull-downs and column-

based assays with the remaining sequences will lend further binding abilities to an expanded reagent set. Of important note is the observation that Illumina-determined copy number, MYcroarray-based fluorescence, and pull-down quality checks all indicated different information about the protein-binding abilities of the sequences examined. An integrative approach that considers each of these techniques, compares across platforms, and incorporates the best of what each has to offer is the ideal in high-throughput and proteome-wide applications such as our own. This story has served as an introductory paragraph on a new page of novel platforms for biologics characterization, and much is yet to be written. For a start, future extensions of this assay include the development of defined reagent sets against HCPs expressed alongside other therapeutic molecules or by other organisms.

Experimental Section

A. HCP production culture

A null CHO host cell line was grown using standard bioreactor conditions at the 50 L scale and harvested by continuous centrifugation. The resulting harvest was diafiltered into PBS, aliquoted and stored at -80 ºC until needed.

B. Selection conditions for M-SELEX (R1, R2) and PD (R3, R4, R5)

Initially, two rounds of M-SELEX with the target on magnetic beads were performed in order to reduce the library diversity before PD screening. Biotinylated aptamers were immobilized on streptavidin beads and resuspended at 10 mg/mL. For the first round of the microfluidic SELEX, 1 nmol of library (6×10^{14} molecules) was incubated with 10 µL of HCP (100 µg target) for five hours before washing. The beads were washed 3 times in

PBSMCT buffer (1x phosphate buffered saline, 1.5 mM $MgCl_2$, and 1 mM $CaCl_2$, 0.01% Tween 20) (Hyclone). The DNA was eluted from the protein target by heating the beads in 50 ul of water for 10 minutes at 95°C. Following elution of the ssDNA off of the beads, the DNA was amplified and purified (Qiagen) yielding dsDNA free of primers. Single-strand generation by lambda exonuclease (New England Biolabs) digestion[19] resulted in DNA for the second round of traditional SELEX.

DNA was immobilized on beads using emulsion PCR to generate monoclonal aptamer particles for PD screening. Aptamer pools on beads were incubated with biotinylated HCPs (bHCPs) for 1.5 hours. The bHCPs that were bound to aptamer particles were labeled with streptavidin phycoerythrin (SA-PE) for 15 minutes so that aptamer particles that were binding to proteins could emit fluorescence and be sorted. Using a BD FACS Aria II instrument, particles with significantly higher fluorescence than background levels were sorted into a collection tube. Particles were sorted at a rate of approximately 2000 events/s for a total of one or more hours. Following sorting, particles were directly amplified with a pilot PCR reaction to determine an optimal cycle number for amplification of all the particles that were collected. The dsDNA products resulting from this full scale amplification were purified using a Qiagen MinElute kit, then amplified via emulsion PCR to begin the next round of PD.

C. Microarray printing

Amino-modified DNA solutions were prepared at 50 uM concentration in Micro Spotting Solution Plus buffer (Arrayit) for coupling on SuperEpoxy 2.0 slides (Arrayit). A

Spotbot 2 (Arrayit) was used to print onto substrates with the following parameters: spacing-250 μm, printing time-0.75 seconds, humidity-80%.

D. Printed microarray assays

Printed slides were blocked for 1.5 hours with solution of 10 mg/mL casein, 0.1% Tween-20, 1 mM $MgCl_2$, in 1x PBS. Binding was carried out with 10 nM biotinylated target (bHCP or bIgG), 2 mg/mL salmon sperm DNA, and PBSMCT with 0.01% Tween-20 for 3 hours. Labeling of biotinylated target bound to slides done with 10 nM Streptavidin-Alexa Fluor 647 conjugate (Life Technologies, Thermo Fisher) in PBSMCT for 5 minutes. Slides were washed with PBSMCT (0.01% Tween-20), PBSMCT with Tween-20 decreased to 0.0025%, PBSMC, and distilled H_2O. Finally, slides were scanned with the red laser and standard red filter on a GenePix 4400A microarray scanner at 2.5 μm resolution.

E. High-throughput sequencing

High throughput sequencing was performed by GeneWiz on an Illumina MiSeq instrument generating approximately 7 million reads for the pool. Of the 7 million reads, approximately 1 million of the reads were unique sequences, and 10% of all the unique sequences were present in 10 copies or higher.

F. Microarray characterization

The sequences obtained from high throughput sequencing were put onto microarrays using *in situ* synthesis (MYcroarray). The array takes a 3 x 20k format, in which the top 6,000 sequences of R3 were synthesized in triplicate in each of the three identical blocks. A

binding assay was carried out with 100 nM HCP through steps of blocking, binding, and washing. Slide was blocked with 0.5% acetylated BSA (AcBSA)/fetal bovine serum (FBS) for 4.5 hours. After blocking, slide was washed once with PBSMCT. Each block was incubated with HCP in a buffer of PBSMCT with 2 µL/mL AcBSA/salmon sperm DNA for 7.5 hours. Following washing of the slide with PBSMCT buffer, the slide was labeled with 10 nM SA-PE for 40 minutes. After a final wash series in PBSMCT, PBSMC, and distilled water, the entire slide was scanned with a GenePix 4400B microarray scanner using the red laser and at a resolution of 2.5 µM.

G. Pull-downs

For DNA-based capture of HCP on beads, MyOne C1 carboxylic acid beads were washed once with PBSMCT, twice with 20 mM NaOH, and three times with PBSMCT. Beads were then incubated in blocking buffer (1% BSA, 1% casein, and 0.1% Tween-20 in PBS) for one hour with end-over-end rotation. After three washes in PBSMCT, 10 µg beads/µg target were combined with biotinylated ssDNA and HCP in binding buffer (PBSMCT with 4 µg salmon sperm DNA/µg target) for two hours with end-over-end rotation. Captured proteins were eluted in denaturing buffer (20% glycerol, 20 mM EDTA, and 50 mM DTT in laemmli buffer) at 94 ºC for ten minutes.

H. SDS-PAGE and 2D DIGE

40 µL of protein samples in denaturing buffer were loaded onto 4-15% TGX pre-cast gels (Bio-Rad cat #456-1084). Gel electrophoresis was carried out at 125 V for 60 minutes. Visualization of bands was achieved after SYPRO Ruby Protein Gel Stain (Invitrogen cat

#S-12000) or SilverQuest Silver Staining (Invitrogen cat #LC6070) using the Gel Doc EZ system (Bio-Rad cat #170-8270).

Applied Biomics (Hayward, CA) performed 2D-DIGE and subsequent spot counting.

I. IP and LC-MS/MS

For affinity-based column chromatography by MedImmune, a 1 mL column of NHS sepharose was coupled with 3'-amino modified 50-mer aptamers. An AKTA system (GE Healthcare) was used to load HCP diluted in PBSMCT+salmon sperm DNA over the column, wash with PBSMCT+salmon sperm DNA, elute with 5 mL PBSMCT+2 M NaCl. At this point, the elutions were TCA precipitated to concentrate proteins.

IP profiling was performed by MS Bioworks to compare all proteins present in an immunoprecipitation sample with proteins present in a control sample. Samples were separated on 10% Bis-Tris Novex minigel (Invitrogen) using MES buffer system. Gels were stained with Coomassie Blue and each lane excised into ten equally sized segments. Gel pieces were processed using ProGest (DigiLab) by: washing with ammonium bicarbonate, washing with acetonitrile, reducing with 10 mM DTT at 60 °C, alkylating with iodoacetamide at RT, digesting with trypsin (Promega) at 37 °C, quenching with formic acid. Supernatant used directly in nano LC/MS/MS with a NanoAcquity HPLC (Waters) system interfaced to a Q Eactive (ThermoFisher). Loaded was on a trapping column packed with Jupiter Proteo resin (Phenomenex), eluted at 350 nL/min. MS and MS/MS at 70k FWHM and 17.5k FWHM resolution, respectively. The 15 most abundant ions selected for MS/MS. Data was searched using Mascot with the following parameters: enzyme-trypsin, database-Uniprot CHO. DAT files were parsed into Scaffold software for validation,

filtering, and to create non-redundant lists per each sample. A total of 255 CHO proteins were detected across both samples with two or more unique peptides: 158 proteins in R5-31, 129 proteins in library, and 186 proteins in R5-150. 25 proteins were unique or 4x higher in R5-31 sample than library. 52 proteins were unique or 4x higher in R5-150 sample than library. Cutoff of 50 spectral counts was used for further analysis of individual HCPs.

References

[1] T. Lai, Y. Yang, S. K. Ng, *Pharmaceuticals* **2013**, *6*, 579–603.

[2] D. C. Krawitz, W. Forrest, G. T. Moreno, J. Kittleson, K. M. Champion, *Proteomics* **2006**, *6*, 94–110.

[3] H. Dorai, S. Ganguly, *Curr. Opin. Biotechnol.* **2014**, *30*, 198–204.

[4] J. Zhu-Shimoni, C. Yu, J. Nishihara, R. M. Wong, F. Gunawan, M. Lin, D. Krawitz, P. Liu, W. Sandoval, M. Vanderlaan, *Biotechnol. Bioeng.* **2014**, *111*, 2367–2379.

[5] X. Wang, A. K. Hunter, N. M. Mozier, *Biotechnol. Bioeng.* **2009**, *103*, 446–458.

[6] M. Cho, S. Soo Oh, J. Nie, R. Stewart, M. Eisenstein, J. Chambers, J. D. Marth, F. Walker, J. a Thomson, H. T. Soh, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18460–5.

[7] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, et al., *PLoS One* **2010**, *5*, DOI 10.1371/journal.pone.0015004.

[8] I. H. Yuk, J. Nishihara, D. Walker Jr, E. Huang, F. Gunawan, J. Subramanian, A. F. J. Pynn, X. C. Yu, J. Zhu-Shimoni, M. Vanderlaan, et al., *Biotechnol. Bioeng.* **2015**, *9999*, n/a–n/a.

[9] J. Godovac-Zimmermann, L. R. Brown, *Mass Spectrom. Rev.* **2001**, *20*, 1–57.

[10] S. Fitter, R. James, *J. Biol. Chem.* **2005**, *280*, 34193–34201.

[11] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 183–187.

[12] C. Tuerk, L. Gold, *Science (80-.).* **1990**, *249*, 505–510.

[13] J. R. Qian, X. H. Lou, Y. T. Zhang, Y. Xiao, H. T. Soh, *Anal. Chem.* **2009**, *81*, 5490–5495.

[14] X. Lou, J. Qian, Y. Xiao, L. Viel, A. E. Gerdon, E. T. Lagally, P. Atzberger, T. M. Tarasow, A. J. Heeger, H. T. Soh, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 2989–2994.

[15] J. Wang, Q. Gong, N. Maheshwari, M. Eisenstein, M. L. Arcila, K. S. Kosik, H. T. Soh, *Angew. Chemie - Int. Ed.* **2014**, *53*, 4796–4801.

[16] T. Schütze, B. Wilhelm, N. Greiner, H. Braun, F. Peter, M. Mörl, V. a. Erdmann, H. Lehrach, Z. Konthur, M. Menger, et al., *PLoS One* **2011**, *6*, 1–10.

[17] P. Jiang, S. Meyer, Z. Hou, N. E. Propson, H. T. Soh, J. a Thomson, R. Stewart, *Bioinformatics* **2014**, *30*, 1–3.

[18] K. M. Ahmad, S. S. Oh, S. Kim, F. M. McClellen, Y. Xiao, H. T. Soh, *PLoS One* **2011**, *6*, e27051.

[19] M. Avci-Adali, A. Paul, N. Wilhelm, G. Ziemer, H. P. Wendel, *Molecules* **2010**, *15*, 1–11.

Chapter V

Summary and future directions

To make something great, we must break apart from the pack. Seeing a connection through to its potential breathes new life into established machines and methods. Mobile phone cameras aren't solely implemented for the purpose of selfies; HTS can't be pigeonholed as a genomics technology only suited to million-base pair reads. It is necessary to head out of the regulated thinking in a laboratory to see that in the hustle and bustle of rural health clinics, workers will rely on more on small, fast, and portable analytical tools such as the P3 and smaRT-LAMP telemedicine platforms. The crossover of aptamer affinity reagents and therapeutic proteins is similarly tricky to approach from the orthodox stance of SELEX. Thinking in terms of convergence and stringency will only lead to an aptamer that is a one-trick pony, whereas the desired result in our screening system is a set of binders that have minimal enrichment, varied structural motifs, and broad diversity with which to capture thousands of crudely processed proteins. The crux of the matter is that these creative strategies work. With the arterial railway of these platforms in place, the next step is to address their limitations and propose how to lay tracks for their successors.

A. Limits, implements, and comments on P3

A big reason that so many care facilities, whether in the first or third world, go without molecular techniques such as PCR is the cost. P3 takes the well-developed technique of PCR and repurposes existing technologies to keep the fixed costs of P3 to a minimal $41.50. Conventional PCR has buildup expenses on the order of $20k—a 500-fold increase above P3. Optimization of equipment goes further than expenses to make P3 practically field-

ready. Carrying out PC-PCR in a laptop instead of a PC would reduce reliance on power sources and increase portability. In true handheld fashion, replacing the UV transilluminator in our work with an add-on illuminator for the phone, consisting of low-cost blue LEDs and blockers to limit ambient light, would provide a repeatable and inexpensive means of end-point readout. Another complicating factor is time. The cycling process of template denaturation and enzymatic amplification requires several hours to complete, but may be expedited with mutant enzymes that are capable of high processivities[1].

Completely demonstrating the diagnostic capabilities of P3 would require thorough testing on a representative set of infected clinical samples. The adversities facing NAT of the gDNA of amastigotes within human blood cells—rather than purified gDNA in samples of human blood—multiply with the difficulties of lysing cells, releasing background human gDNA, and potentially introducing closely related species of human pathogenic trypanosomes. Though a background of thousands of host cells to a few copies of target gDNA is hardly trivial, one of the earliest examples of camera-recorded PCR showed amplification of a 142-bp product from 50 ng of the HIV genome even in 40,000 cell equivalents of human gDNA[2]. Modern advances have improved amplification abilities with inhibitor-resistant chemistries and specificity-boosting additives[3], making us optimistic about the possibility of reducing off-priming to unrelated gDNA in blood obtained from patients with Chagas disease. With ample genetic information available, this system can even move beyond the singly demonstrated application for Chagas disease. PCR has been applied to a vast range of infectious diseases, meaning that the P3 system could diagnose microbial pathogens responsible for HIV, hepatitis, and tuberculosis, to name a few.

B. Limits, implements, and comments on smaRT-LAMP

The performance of computer processing has advanced according to Moore's Law over the past four decades so that electronic devices continue to become smaller, cheaper, and faster[4,5]. In the face of these advancements, we can engineer handheld diagnostic tests such as smaRT-LAMP that could truly put into action by many users in various settings. Future work with this platform is only limited by our knowledge of target genomes. Reagents and primers for performing LAMP on West Nile Virus and Koi herpesvirus are readily available in a kit-based format (Eiken Chemical Co., Ltd.), making these two immediately available scenarios for future testing in low-resource settings.

This union of rapid DNA amplification and real-time test readout is suited for work in the field, meaning that smaRT-LAMP could be physically improved by modifying the reaction setup to be entirely transportable. The blue light to illuminate the fluorescent indicator FDR and visualize the progress of qRT-LAMP could very well be emitted from the smartphone as another function of the Bacticount app; likewise, the blackout box used to block ambient light during the reactions could take the form of a clip-on module for the phone. Finally, providing heat is as simple and inexpensive as combining calcium oxide and water[6]. With these modifications, our prototype could be transformed into an instrument-free methodology that is wholly operable outside of labs or clinics.

C. Limits, implements, and comments on proteomic QPASS

To quote Sumedha Jayasena's words of fast-paced frisson on aptamers, they "rival antibodies in both therapeutic and diagnostic applications." An expansion on this is that antibody-based detection methods face competition from aptamers as alternative affinity

reagents. Antibodies are the affinity reagent of choice for picking out HCPs from biotechnology products in ELISAs, but the use of a sandwich assay for characterization is hugely different than the use of resin-bound pAbs for the actual removal of impurities from a product. The development of pAbs suffers greatly from variations between batches, and the pAbs used to determine HCP levels in a drug will most likely be those that will give the best results for FDA approval in an ELISA or Western blot rather than the most accurate measurement of purity. Our defined set of aptamer reagents could work against this trend of making the ends justify the means. As bioinformatics is continuously pushing big data to the extreme, the time is right to seek out different modes and means of quantifying HCP coverage. We demonstrated via the QPASS platform that our aptamers bind to HCPs when immobilized on arrays—and the easy readout format of DNA microarrays is where future directions in deconvoluting the CHO proteome should focus. There are issues to overcome with attaining extremely sensitive limits of detection for HCP targets in highly convoluted background mixtures, as witnessed by our own specificity tests with IgG and HCPs, but an appropriate blocking or decoupling strategy could turn printed and *in situ* microarrays into the next standardized tool for industry testing of contaminants in recombinant therapeutics.

D. Outlook

Applications of genetic testing have spurred the creation of many innovative systems—but creativity is not enough to foster results. In terms of NAT, isothermal techniques have been differentiated from PCR with claims of the former having more pros and fewer cons (*i.e.*, increased speed to completion and reduced need for equipment). Early reports in characterizing the tolerance of LAMP for biological substances implied that LAMP was more robust and sensitive than PCR[7–9]—though contrary results now warn against broad

generalizations[10,11]. As we progressed through the characterization of our smaRT-LAMP and P3 systems, we heeded these cautionary tales: in each case with purified, crude, and animal samples, we took care to record the actual, observed trends rather than what might be expected from the many precedents set with PCR. Prior researchers have neglected to differentiate reactions such as rolling circle amplification (RCA), strand displacement amplification (SDA), or LAMP from PCR when carrying out classical protocols developed for nucleic acid extraction, signal measurement, or post-amplification analysis[12]. This desire to broadly lump all DNA amplification technologies together is at odds with their fundamental differences. Practicality demands that there must be an empirical dimension to verifying NAT by experimental optimization.

Today's diagnostics engineers are facing issues beyond the technological; theirs is an uphill battle against societal norms. Consider our own nation, where the ineffective monitoring and processing of food continually results in mortalities from outbreaks of foodborne infection. Poultry, swine, and cattle farmers regularly incorporate growth-producing antibiotics into animal feeds, which greatly increase antimicrobial resistance while weakening the animals' own resistance. Unnecessary distribution of antimicrobial agents in both human and veterinary medicine has the same effect[13]. Developing more sensitive and rapid molecular methods to detect enteropathogens in food can only accomplish so much when preventative and preemptive efforts are brushed aside with a cavalier disregard. Here, we mobilize genetic testing with devices that are advancing exponentially. The next step is bringing about an equally fast-paced revolution with healthcare. Considering that a reinvention of internal regulation is required in one of the world's most technically advanced countries, it should come as no surprise that third-world

nations affected by neglected tropical diseases are in dire need of some investment and advocacy to reformulate their health systems. Very few resources are invested in fighting the global threat of Chagas disease[14]. Movements to coordinate resources for controlling disease transmission and providing access to treatment require a strong partnership between endemic communities and relief sources.

NAATs have been heralded as a full-force gale bringing the winds of change to POC diagnostics, but this promise has yet to bloom into fruition. Perhaps the most salient example is that it has taken nearly 20 years since the development of PCR for a molecular diagnostic test to gain approval for use outside of typical hospital and physician settings. In January of this past year, Alere's i Influenza A&B test received the Clinical Laboratory Improvement Amendments (CLIA) waiver for broad usage by healthcare providers[15]. Despite the general lack of commercially available POC diagnostics that offer both sensitivity and accuracy, the recent achievement by Alere Inc. can serve as a guiding light towards a path for deployable NAT technologies. Our hope would be that DNA-based tests will continue to spin like a hurricane into the business of biotesting. The three applications presented in this thesis—transforming PCs for Chagas disease detection, profiling salmonellae biomarkers in animal samples with phones, and detecting impurities in biopharmaceuticals with aptamers—have electrified the space between different technologies by merging them together. DNA's dual nature as a protein-binding entity and a helical unit truly do provide it with two twin energies that can fuel our generation's advancement into the brave new world of healthcare.

References

[1] W. M. Barnes, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 2216–2220.

[2] R. Higuchi, C. Fockler, G. Dollinger, R. Watson, *Biotechnology. (N. Y).* **1993**, *11*, 1026–1030.

[3] A. Trombley Hall, A. McKay Zovanyi, D. R. Christensen, J. W. Koehler, T. Devins Minogue, *PLoS One* **2013**, *8*, e73845.

[4] A. Ozcan, *Lab Chip* **2014**, DOI 10.1039/c4lc00010b.

[5] G. Miller, *Perspect. Psychol. Sci.* **2012**, *7*, 221–237.

[6] P. LaBarre, K. R. Hawkins, J. Gerlach, J. Wilmoth, A. Beddoe, J. Singleton, D. Boyle, B. Weigl, *PLoS One* **2011**, *6*, e19738.

[7] P. Francois, M. Tangomo, J. Hibbs, E.-J. Bonetti, C. C. Boehme, T. Notomi, M. D. Perkins, J. Schrenzel, *FEMS Immunol. Med. Microbiol.* **2011**, *62*, 41–8.

[8] K. a. Curtis, D. L. Rudolph, S. M. Owen, *J. Virol. Methods* **2008**, *151*, 264–270.

[9] H. Kaneko, T. Kawana, E. Fukushima, T. Suzutani, *J. Biochem. Biophys. Methods* **2007**, *70*, 499–501.

[10] G. Nixon, J. A. Garson, P. Grant, E. Nastouli, C. A. Foy, J. F. Huggett, *Anal. Chem.* **2014**, *86*, 4387–4394.

[11] G. Kiddle, P. Hardinge, N. Buttigieg, O. Gandelman, C. Pereira, C. J. McElgunn, M. Rizzoli, R. Jackson, N. Appleton, C. Moore, et al., *BMC Biotechnol.* **2012**, *12*, 15.

[12] A. Niemz, T. M. Ferguson, D. S. Boyle, *Trends Biotechnol.* **2011**, *29*, 240–50.

[13] D. G. Maki, *N. Engl. J. Med.* **2009**, *360*, 949–953.

[14] J. Gascon, R. Vilasanjuan, A. Lucas, *Expert Rev. Anti. Infect. Ther.* **2014**, *12*, 393–5.

[15] *Clin. Lab Prod.* **2015**.